UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

GENETIC AND GENOMIC ANALYSIS OF BAHD ACYLTRANSFERASES THAT DECORATE CELL WALL COMPONENTS WITH PHENOLIC ESTERS AND ALTER PLANT BIOMASS RECALCITRANCE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHENGCHENG ZHANG
Norman, Oklahoma
2018

GENETIC AND GENOMIC ANALYSIS OF BAHD ACYLTRANSFERASES THAT DECORATE CELL WALL COMPONENTS WITH PHENOLIC ESTERS AND ALTER PLANT BIOMASS RECALCITRANCE

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

Dr. Laura Bartley, Chair

Dr. Anthony Burgett

Dr. Scott Russell

Dr. Marc Libault

Dr. Ben Holt

Acknowledgements

There are many people without whom this dissertation would not have been possible. My deepest gratitude is to my advisor, Dr. Laura Bartley. During my graduate training with her, I gained the ability to learn actively and think critically, which will benefit my entire life. She is such an encouraging and supportive professor, who always guides me out of confusions and difficulties. She genuinely cares about her students and believes that we can succeed. With her help and encouragement, I overcame many difficulties and gradually gained self-confidence to be a woman scientist. She is also thoughtful especially when I was pregnant and in the postpartum period. I am always feeling lucky to have her as my advisor.

I also would like to acknowledge the support of my committee members and collaborators. I want to thank Dr. Kessler, Dr. Burgett, Dr. Russell, and Dr. Libault for all valuable suggestions on my graduate proposal. Thank Dr. Kessler and Dr. Holt for sharing *Arabidopsis* growth chamber and all their technical support. I want to thank Dr. Masly's course, R programming, which benefits my research a lot. Thank Dr. Jizhong Zhou for sharing the server so that I could fulfill the RNA-seq analysis. Thank Dr. Steve Karlen and Dr. Rebecca Smith from the Ralph lab at University of Wisconsin-Madison and Dr. Alex Tsai from the Scheller lab at Joint BioEnergy Institute for helpful discussions and wonderful collaborations.

I am grateful to all my old and current colleagues, Dr. Matt Peck, Sandra Thibivilliers, Dr. Kangmei Zhao, Dr. Fan Lin, Mary LaPorte, Bisoh Moma, David Thomas, Dr. Daniel Jones, Dr. Zhenzhen Qiao, Dr. Chamindika Siriwardana, Dr. Prince

Zogli, Dr. Swadhin Swain, Hengping Xu, and Jing Yuan. I enjoyed working with you and learned a lot from you. Thank you for all your support and understanding.

Special thanks go to my dear parents for their continuous support and trust. Thank my mother-in-law for her kind caring of my little girl. Most importantly, I want to thank my dear husband, Tao Xu, who is my best partner in both science and life. Without him, I would not have today's achievement and my beautiful baby. Last, I want to thank my little girl, Mia Xu. You give me smiles every day and make my life full of sunshine.

Table of Contents

List of Tables

List of Figures

Abstract

Next generation biofuels make use of the energy stored in plant cell walls, so called lignocellulosic biomass. However, the natural resistance of plant cell walls against deconstruction, i.e., "cell wall recalcitrance", poses a significant challenge to large-scale commercialization of lignocellulosic biofuels. Grasses, including cereal crops and perennial grasses, are the major source of terrestrial biomass. Previous studies have implicated phenolic acids (ferulic acid and *p*-coumaric acid) in grass cell wall recalcitrance, but the enzymes that attach these molecules to cell wall precursors are largely unknown. To address this gap, this dissertation accomplished the following: 1) genetic characterization of two so-called "BAHD" acyltransferase genes from rice (*Oryza sativa, Os*), *OsAT5* and *OsAT9*; 2) examination of the impact of feruloylated lignin on biomass recalcitrance; 3) exploration of cell wall recalcitrance genes in the bioenergy crop, switchgrass, via transcriptomics.

To study the enzymes involved in FA decoration in cell walls, we overexpressed a BAHD -coenzyme A acyltransferase, encoded by the *OsAT5* gene, in rice and found increased incorporation of feruloyl monolignol conjugates (ML-FAs) in lignin. Cell wall chemical and gene sequence phylogenetic analysis of gymnosperms, dicotyledonous and monocotyledonous plants revealed that incorporation of ML-FAs is wide-spread in angiosperms, but those orthologous genes to *OsAT5* are only present in grasses and other commelinid monocot species. These results suggest that angiosperms have convergently evolved the ability to synthesize this newly recognized conjugated lignin precursor.

To verify the enzymatic activity and the role of OsAT5 in cell wall decoration, we heterologously expressed the *OsAT5* gene in yeast and *Arabidopsis*, which naturally

lack monolignol ferulates. Contrasting the cell wall properties of wild type and transgenic *OsAT5* overexpression-rice (*Ubi*$_{pro}$-*OsAT5*) and -*Arabidopsis* (*C4H*$_{pro}$-*OsAT5*) revealed that, 1) transgenic *Arabidopsis* exhibited reduced cell wall recalcitrance, whereas transgenic rice lines did not; 2) compared with transgenic rice, transgenic *Arabidopsis* has a more significant impact on sinapyl ferulate (S-FA) incorporation; 3) when treated with alkaline, wild-type rice lignin, which naturally possesses ML-FA conjugates, shows a higher solubility than wild-type *Arabidopsis* lignin, consistent with grasses having a more chemically-labile lignin polymer than dicots. The discordant observations in rice and *Arabidopsis* indicate that ML-FAs produced by OsAT5 have differential impacts on cell wall traits depending on the plant species and tissue, raising the possibility of tailoring lignin structure engineering to different species to tune biomass recalcitrance.

To reveal other mechanisms of feruloylation in grasses, we genetically characterized another BAHD acyltransferase, *OsAT9,* in rice. Overexpression of *OsAT9* in rice with the maize *Ubi* promoter increased the ratio of ferulic acid to *p*-coumaric acid (FA:*p*CA) in cell wall polysaccharides and improved extractability of xylan with base treatment, but reduced the enzymatic digestibility of the leaf and stem. These results suggest that OsAT9 is a strong candidate as a feruloyl arabinosyl transferases responsible for feruloylation of rice arabinoxylan by which biomass recalcitrance can be altered.

The important agricultural and industrial use of the perennial grass, switchgrass, has generated particular interest in dissecting biomass digestibility-related genes. To accomplish this, we conducted RNA sequencing of four switchgrass genotypes with distinct digestibility, including four sample types (whole elongation 4-stage tiller, leaf, soft stem, and hard stem). The transcriptomes allowed dissection of tissue-specific, lignin

biosynthesis- and biomass digestibility-associated genes. We discovered that some protein kinases and cell wall biosynthesis genes are highly related to biomass digestibility and also noted subfunctionalization of putative cell wall-decorating BAHD acyltransferase and lignin biosynthesis genes.

This dissertation significantly expands knowledge of cell wall decoration by ferulates, provides insight into the functions of BAHD acyltransferase gene family members and their impacts on cell wall synthesis and biomass recalcitrance in both model plants and food and energy crop species. This study also provides valuable information and new ideas for plant breeding and engineering to create less recalcitrant plant biomass for industrial use and animal forage.

Keywords: Biofuel, lignocellulose, rice, switchgrass, grass, dicot, lignin, cell wall, ferulic acid, BAHD acyltransferases, transcriptome

Chapter 1 Introduction

1.1 Motivation

Plant cell walls contain the three most abundant carbonaceous polymers in the world-cellulose, hemicellulose, and lignin. Grasses are promising feedstocks for the production of next-generation fuels that can replace fossil fuels and meanwhile reduce greenhouse gas emissions. However, recalcitrance of biomass cell walls to depolymerization poses a significant challenge to large-scale commercialization of lignocellulosic biofuels production. It is known that the intrinsic structural features of major cell wall polymers (cellulose, hemicellulose, and lignin) and cross-linking structures formed between these polymers, not only confer plants with mechanical strength and resistance to environmental stress, but also fortify cell wall recalcitrance during bioconversion. However, knowledge of the mechanism of how plants construct their cell walls is still very limited. Ferulic acid, a phenylpropanoid, extensively decorates grass cell wall polymers and is considered as an important crosslinking agent in grasses. Encouragingly, a few studies on the BAHD acyltransferase gene family pointed out strong candidates that might be responsible for cell wall cross-linking. Much effort still needs to illustrate their functions in cell wall decoration and how these decorations affect cell wall recalcitrance. Beyond that, the diversity of gene family members varies among plants as a result of evolution (for example, diploids versus polyploids), raising questions about their specific biological functions across plants, genotypes, tissues, developmental stages, as well as cell types. Therefore, understanding of relevant mechanisms in both tractable plants, such as rice, and less tractable but important bioenergy crops, such as switchgrass, will hopefully fill in our knowledge gap and provide insights on strategies for plant cell

wall engineering to produce easily digestible biomass. Below, I will introduce the background about current status of biomass utilization, reported cell wall recalcitrant factors, genetic evidence for the BAHD gene family in cell wall decoration and recalcitrance, and conclude with the specific aims of this dissertation.

1.2 Lignocellulosic feedstocks

Compared with current commercial biofuels made from sugars, starch, and oils in food commodities, next-generation biofuels use nonfood biomass with the potential of improving energy yields and reducing greenhouse gas emissions (Farrell et al., 2006; Fargione et al., 2008; Schmer et al., 2008). Polysaccharides (cellulose and hemicellulose) and lignin in terrestrial plant cell walls are and abundant and possibly sustainable chemical energy that can be converted to biofuels. This so-called lignocellulosic material includes biomass from dedicated energy crops; agricultural and industrial waste products, such as corn stover, wheat straw; and paper mill waste. Apart from well-known ecological importance (stabilizing soils and minimizing soil erosion), grasses have a high biomass yield, providing ~57% of the biomass used for biofuels production in the U.S. (Perlack et al., 2005). Switchgrass, Sorghum, and Miscanthus are being developed as herbaceous energy crops by the U.S. Department of Energy (U.S. DOE, 2016). Notably, switchgrass can produce 5.2-11.1 Mg of dry biomass per hectare in field trials even on marginal cropland in the mid-continental United States (Schmer et al., 2008), and various switchgrass cultivars have been found able to adapt to diverse environmental conditions across North America (Sanderson et al., 2006).

Due to the polyploid nature of switchgrass, grasses with a simpler genome are needed to facilitate fundamental studies of the molecular mechanisms of cell wall construction. Rice

(*Oryza sativa*), Brachypodium (*Brachypodium distachyon*) and foxtail millet (*Setaria viridis*) are diploid grasses with smaller genomes (382 Mbp, 272 Mbp, and 515 Mbp, respectively) and that are relatively easily grown, thus serving as useful models for genetic and biochemical/chemical studies (Doust et al., 2009; Bevan et al., 2010). Among them, rice is more favored in this study because: 1) rice straw alone composes 23% of the total crop residue (Lal, 2005); 2) it has a well-assembled genome and rich genomic resources already developed, including gene expression database and mutant resources (Krishnan et al., 2009). Studies on rice cell walls can give important clues for understanding and/or engineering other grasses.

1.3 Biochemical conversion of lignocellulosic biomass

Biochemical conversion and thermochemical conversion are two major approaches for biofuel production from lignocellulosic biomass; the former is relatively closer to industrial implementation. Biochemical conversion is also known as direct microbial conversion or biological conversion, which consists of three major steps: pretreatment, saccharification, and fuel synthesis. In the first step, harvested biomass is chopped and pretreated to destroy microstructure and improve the accessibility of polysaccharides. Conventional pretreatments include a combination of heat, pressure, acid, and/or base treatment (Agbor et al., 2011); use and recycling of ionic liquids is a newer and highly effective pretreatment (Li et al., 2010a). In the second step, enzymes are added to the neutralized slurry to breakdown cellulose and other polysaccharides into simple sugars, a process known as saccharification. Finally, microbial fermentation is engaged to convert resulting sugars into fuels. Although a diversity of biofuels (e.g., ethanol, butanol, alkanes, and biodiesel) can be produced in this way (Peralta-Yahya and Keasling, 2010),

the cost and inefficiency of converting polysaccharides to fermentable sugars are challenged by cell wall recalcitrance impeding commercialization of lignocellulosic biofuels production (Lynd et al., 2008).

1.4 Effect of grass cell wall polymer features on biomass recalcitrance

Cell wall composition, polymer features, and interactions among polymers all contribute to cell wall recalcitrance. Vascular plants have a similar architectural construction in cell wall. They mainly have cellulose microfibril as the scaffold, which is embedded in cell wall matrix built up with a large proportion of non-cellulosic polysaccharides, a small number of structural proteins and glycoproteins, as well as phenolic components. There are two types of cell wall: primary cell walls surround enlarging cells, whereas secondary cell walls located on the inner side of primary walls restrict cells from expansion (Cosgrove and Jarvis, 2012). Cell wall composition differs among plant species as well as primary and secondary cell walls. In this part, I will focus on the features of three major wall components (cellulose, hemicelluloses, and lignin) and crosslinking structures that have been characterized as cell wall recalcitrance factors in grass.

1.4.1 Cellulose

Cellulose is the predominant polymer in grass cell walls, accounting for 37-43% of dry matters of switchgrass (Sarath et al., 2007; Vogel et al., 2011) and 21-45% in rice biomass (Tanaka et al., 2003; Johar et al., 2012). It also makes up about 30-40% of the dry mass of both primary and secondary cell walls (Cosgrove and Jarvis, 2012). Cellulose consists of long chains of 500 to 15,000 ß-(1-4)-covalently bonded glucose residues; hydrogen bonds between ~36 chains form compact crystalline microfibrils that exclude internal water solvation and form crystalline surfaces that are inaccessible to enzymatic hydrolysis

(Somerville, 2006). From the perspective of its use in biofuels production, cellulose hydrolysis provides more abundant and fermentable C6-sugars for biofuel-producing microbes.

Cellulose crystallinity and cellulose polymerization degree (DP) are two major features of cellulose but negatively correlated with enzyme-based biomass digestion after base and acid pretreatments (Wang et al., 2016). Studies showed that cellulose DP is positively correlated with crystallinity index and the reduction in DP significantly improved biomass digestibility. However, it is still a controversy on the role of DP in lignocellulose recalcitrance.

1.4.2 Xylan

Grass xylan is the most abundant hemicellulose, with complex linkages and sidechain structures. Relative to dicots, grasses have a very high abundance of xylan in both primary and secondary cell walls. Xylan accounts for ~50% of grass hemicellulose (Scheller and Ulvskov, 2010). Approximately 20-25% of dry switchgrass biomass is xylan (David and Ragauskas, 2010). Biochemically, xylan consists of a backbone of ß-(1-4) linked xylose residues and various substitutions (**decorations**) in grasses (**Figure 1**). Principally, grass xylose is periodically linked at the O-3 or O-2 position to the 1-carbon of arabinose residues to form arabinoxylan (AX). The arabinose residues are in the furanose (f) form with five atoms in the sugar ring (Araf), rather than in the pyranose form (p), which has a 6-membered ring. Some of arabinose residues in xylan are modified at the 5-carbon by acylation with hydroxycinnamic acids, especially, ferulic acid (FA) and to a lesser extent, *p*-coumaric acid (*p*CA) (de O Buanafina, 2009; Bartley et al., 2013). Besides, grass arabinoxylan contains some minor substitutions on Araf, such as β-(1->2)Xyl-(1->2)Gal

(Saulnier et al., 1995), β-(1->2)-Gal and β-(1->2)-Xyl (Wende and Fry, 1997; Chiniquy et al., 2012) and substitutions shared with dicot secondary cell walls, such as (4-O-methyl-) glucuronosyl at the O2- position (Glucuroarabinoxylan, GAX) and acetylation (Scheller and Ulvskov, 2010).

The complexity of grass xylan sidechains influences the water solubility of xylan and its interaction with cellulose and lignin, therefore requiring a suite of expensive hydrolytic enzymes to cooperatively breakdown biomass. Recently, studies found that the pattern of xylan decoration by acetate or glucuronic acid is maintained among vascular plants and is critical for interaction of xylan with hydrophilic faces of cellulose fibrils, and negatively related biomass digestibility (Busse-Wicher et al., 2016; Grantham et al., 2017). Increasing Araf substitution on xylan in the mature tissues could enhance the interaction between arabinoxylan and cellulose via hydrogen bonds, affecting cellulose crystallinity (Li et al., 2015). Grasses have evolved intrinsic mechanisms to manipulate the substitution degree of arabinoxyl on xylan. For example, more substitutions are present in young tissues than in mature tissues, which may suggest an arabinofuranosidase-mediated removal mechanism during maturation (Chávez Montes et al., 2008; Lin et al., 2016). In terms of association with biomass digestibility, decreasing substitution on xylan reduces enzymatic degradation of plant biomass to usable sugars. This may be explained by the hypothesis that less substitution may dampen substrate features by which specific hydrolases can efficiently recognize and bind to. Moreover, FA on GAX can undergo radicle oxygen-mediated coupling to form ether bonds or C-C bonds , making diferulates and triferulates that result in xylan-xylan cross-linking (Takahama and Oniki, 1994; Bunzel et al., 2008) (**Figure 1.1**). FA on GAX has also been

proposed to nucleate lignin formation in grasses (Bunzel et al., 2004) to promote xylan-lignin cross-linking, again, impeding the enzyme digestibility of cell walls (Grabber et al., 1998). Studies observed that grass with less feruloylation on GAX is more easily converted into biofuels (Akin, 2008; Piston et al., 2010; Matias de Oliveira et al., 2014). Therefore, modification of xylan sidechains and FA-mediated crosslinking is important in biomass deconstruction.

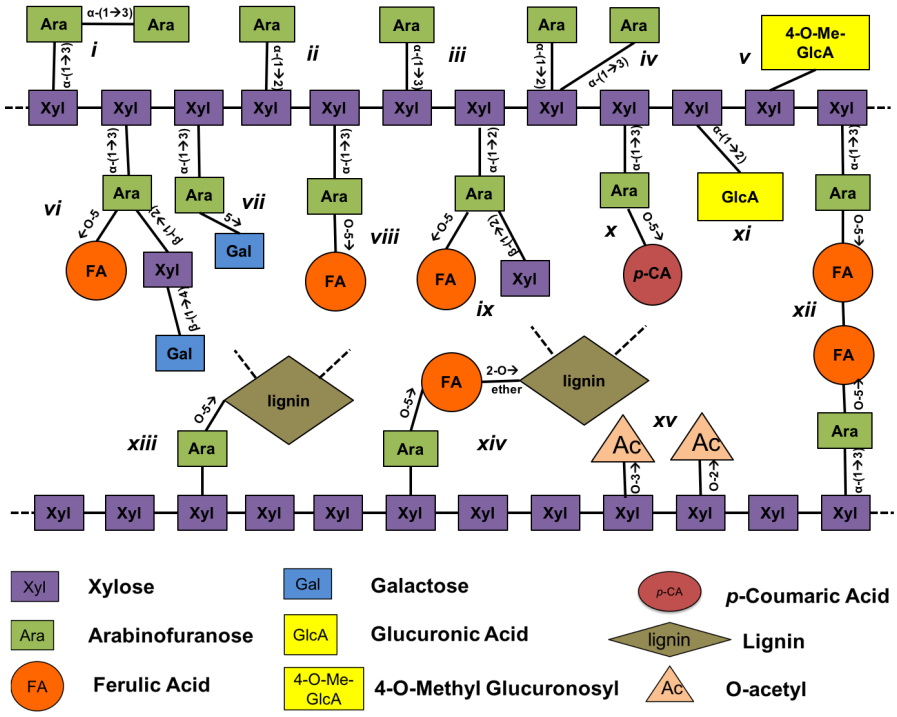


Figure 1.1 Structural features observed on heteroxylans of grasses.
i. Araf-α-(1,3)-Araf-α-(1,3)Xyl; ii. Araf-α-(1,2)-Xyl; iii. Araf-α-(1,3)-Xyl; iv. Araf-α-(1,2)[Araf-α-(1,3)]-Xyl; v. (4-O-Me)GlcA-α-(1,2)-Xyl; vi. Galp-β (1,4)-O-Xyl-β(1,2) [5-O-(trans-feruloyl)]Araf-α-(1,3)-Xyl; vii. [5-O-Galp]Araf-α-(1,3)-Xyl; viii. [5-O-(trans-feruloyl)]Araf-α-(1,3)-Xyl; ix. Xyl-β(1,2)-[5-O-(trans-feruloyl)]Araf-α-(1,3)-Xyl; x. [5-O-(trans-*p*-coumaroyl)Araf-(1,3)-Xyl; xi. GlcA-α-(1,2)-Xyl; xii. Ester linked FA dimer; xiii. Araf-lignin; xiv. FA ether linked to lignin; xv. O-3 and/or -2-acetyl-Xyl.

1.4.3 Lignin

Lignin is a complex phenolic polymer with various structures (**Figure 1.2**). It is deposited in interstices of cell walls during secondary development (Terashima et al., 2004),

providing mechanical support to plant terrestrial life (Vanholme et al., 2010). The lignin content in grasses and woody plants accounts for 15-30% by weight and 30-40% by energy content in the total biomass (Boerjan et al., 2003). In general, lignin negatively affects enzymatic hydrolysis by preventing cellulose microfibril swelling, reducing surface area access for cellulase enzymes and/or preventing cellulase action on cellulose surface (Wang et al., 2016). Genetic studies on lignin biosynthesis in plants found reducing the content of lignin improves biomass digestibility. However, disruption of lignin biosynthesis usually sacrifices plant growth, stress tolerance, and biomass yield, such that directly manipulating lignin generation is less prioritized for plant biomass engineering (Wang et al., 2016).

Lignin biosynthesis relies on generation of three main monomers (*p*-coumaryl alcohol, coniferyl alcohol, and sinayl alcohol) and subsequent radical coupling (Boerjan et al., 2003). There are more than ten enzymes sequentially involved in monolignol biosynthesis, including phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), shikimate hydroxycinnamoyl transferase (HCT), coumarate 3- hydroxylase (C3H), caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), ferulate 5-hydroxylase (F5H) and caffeic acid/5-hydroxyferulic acid O-methyltransferase (COMT). These building blocks are turned to *p*-hydroxyphenyl (H), guaicyl (G), and syringyl (S) subunits once incorporated to lignin polymers. They present four, three, and one potential covalent branch site(s), respectively (Boerjan et al., 2003) such that higher amounts of H and G lignin potentially build a more branched lignin structure. Monolignols (MLs) acylated by phenolic acids (especially *p*CA, FA, and *p*-hydroxybenzoate) and acetate, are

accepted as monomers of lignification in various species. The flavonoid tricin recently discovered in monocot lignin (Del Río et al., 2012) also functions as a nucleation site that initiates lignin polymer chains in maize stover (Lan et al., 2015).

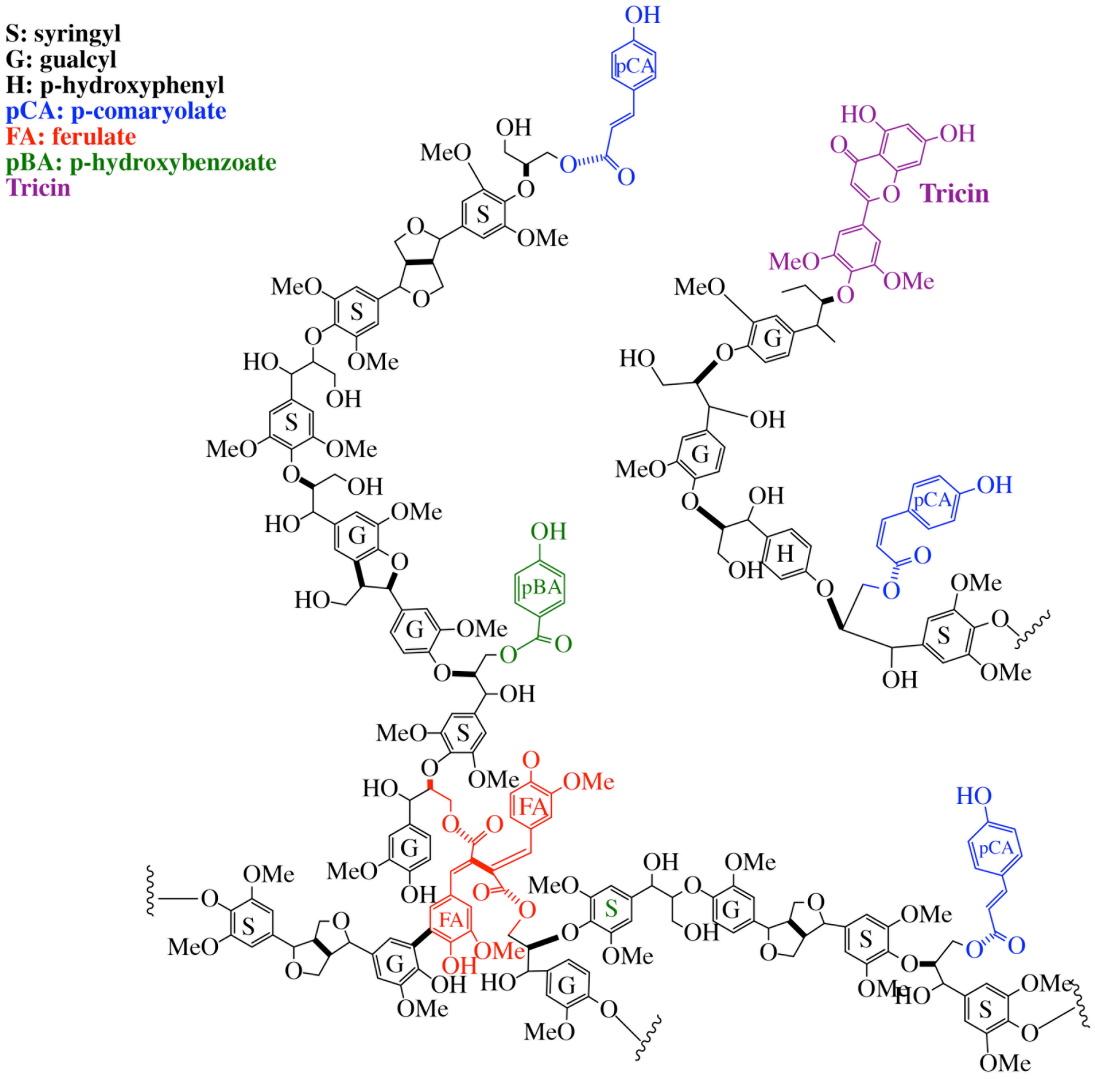


Figure 1.2 Lignin structure.

The features of lignin structure alter cell wall recalcitrance

S:G ratio

Since S and G are predominant lignin types in grasses, the S/G ratio is, to a large extent, indicative of lignin structure. However, the effect of this ratio on biomass

recalcitrance is controversial depending on plant species and pretreatment involved. For example, several studies found higher S:G ratio correlates with less enzymatic hydrolysis of un-pretreated biomass (Tu et al., 2010; Zhang et al., 2011) and NaOH and H_2SO_4 pretreated biomass (Jiang et al., 2016). Researches deduced a negative effect to a more efficient coverage of S-lignin (extended shape) than G-lignin (branching) on cellulose fibrils (Besombes and Mazeau, 2005a; Besombes and Mazeau, 2005b). However, higher S:G ratio correlates with more enzymatic hydrolysis for untreated poplar trees (Studer et al., 2011; Mansfield et al., 2012) and hot water pretreated Arabidopsis (Li et al., 2010b), green liquor and Kraft pretreated Eucalyptus (Papa et al., 2012; Santos et al., 2012). With these discrepant observations, we can speculate that the S:G ratio partially contributes to biomass recalcitrance; the complexity of plant biomass and pretreatment-caused concomitant changes in other cell wall structure may buffer effects arising from lignin composition.

H-lignin

Due to the low occurrence of H-lignin in plants, effect of H-lignin on biomass recalcitrance has been less investigated. Interestingly, studies revealed that high H-lignin plants exhibited reduced biomass recalcitrance. For instance, the H/G ratio in the KOH-extractable lignin of wheat and rice presented a strong positive correlation with enzymatic hydrolysis of pretreated biomass (Wu et al., 2013); the disruption of a transcriptional co-regulatory mediator MED5a/5b in a severe lignin deficient Arabidopsis mutant where C3'H, a key enzyme in the early stage of phenylpropanoid pathway has been reduced, synthesized nearly wild type amount of pure H lignin in Arabidopsis and reduced biomass recalcitrance (Bonawitz et al., 2014). These positive effects of H-lignin on biomass

digestion might be ascribed to reduced lignin molecular weight, reduced cellulose crystallinity via H unit-glucan bonding, and higher linkage activities between H monomers than G and S (Ragauskas et al., 2016).

***p*-coumaryl-lignin**

A characteristic of grass cell wall is that grass monolignols are acylated by *p*-coumaric acid at the γ- carbon of hydroxyl group on the alcohol at the end of the propanoid "carbon tail" (**Figure 1.2**) (Ralph, 2010). *p*-coumaroyl esters may act as "radical catalysts", rapidly passing the radical to sinapyl alcohols, thereby potentially facilitating lignin polymerization (Takahama and Oniki, 1994; Ralph, 2010). Attachment of *p*CA to lignin at this particular location has been found in a diversity of grass species (Soreng et al., 2015) including maize, bromegrass, bamboo, sugarcane, elephant grass, rice (Withers et al., 2012; Karlen et al., 2016; Takeda et al., 2017), switchgrass (Shen et al., 2009), and Brachypodium (Petrik et al., 2014). Recently, *p*-coumaryl lignin structure was also found in other commelinid orders (*Zingiberales, Commelinales, and Arecales*) (Karlen et al., 2018). Biochemical analysis suggests that this modification may indirectly enhance lignin polymerization (Ralph, 2010). *Arabidopsis* and *Brachypodium* genetically engineered with more *p*-coumarates in lignin polymers are more digestible, probably due to more alkaline-solubilized lignin structure (Petrik et al., 2014; Sibout et al., 2016).

Feruloyl lignin (zip-lignin)

Feruloyl lignin contains feruloyl monolignol conjugates (ML-FAs) and it is a promising lignin structure that can facilitate depolymerization of lignin polymers. Previous studies revealed incorporation of synthesized coniferyl ferulate into lignin, which enhanced alkaline delignification and enzymatic hydrolysis (Grabber et al., 2008;

Ralph, 2010). It was also confirmed by introducing a feruloyl-monolignol transferase (FMT) from Chinese angelica [*Angelica sinensis* (*As*), a dicotyledonous Chinese medicinal plant] to polar. Heterologous expression of *AsFMT* in poplar tree (*Populus alba* × *Populus grandidentata*) generated ML-FAs and incorporated into lignin polymers. The resulting biomass presented improved saccharification after mild base pretreatment (Wilkerson et al., 2014). In addition, suppression of the first lignin specific biosynthetic enzyme, cinnamoyl-CoA reductase (CCR) resulted in an increase in the intercellular pool of feruloyl-CoA and in ML-FAs in maize lignin polymer, and a decrease in lignin content and monomers, therefore enhancing the digestibility of stem rind tissue (Smith et al., 2017). The mild alkaline cleavable ML-FA conjugates are a promising target for engineering bioenergy crop with low cost on lignin removal. However, we still do not know if this lignin structure is naturally existing in grasses or other plants, and what enzymes are responsible for producing this conjugate.

Candidate Enzymes for feruloylation and *p*-coumarylation in cell walls

We have gradually gained knowledge on enzymes responsible for the incorporation of *p*CA and FA into grass cell walls. The acids (*p*CA and FA) acylate lignin in the cytosol. And the acylated monolignol conjugates can be transported to cell walls and used as nontraditional monomers for lignification (Ralph, 2010). The BAHD protein family is a large family involved in acylation of secondary metabolites by CoA thioesters in plants and then named by the first four characterized members (**B**EAT or benzylalcohol *O*-acetyltransferase from *Clarkia breweri*; **A**HCTs or anthocyanin *O*-hydroxycinnamoyltransferases from *Petunia, Senecio, Gentiana, Perilla,* and *Lavandula*; **H**CBT or anthranilate *N*-hydroxycinnamoyl benzoyltransferase from *Dianthus*

caryophyllus; **D**AT or decetylvidoline 4-*O*-acetyltransferase from *Catharanthus roseus*) (Bontpart et al., 2015). Site-directed mutagenesis and structural analysis of the first crystalized BAHD member, malonyltransferase Dm3MaT3, found that the conserved motif HXXXD is crucial to its catalytic activity and also presents in other thioester CoA-utilizing acyltransferase families; DFGWG at the C terminal is another important motif, whose variation largely affects enzymatic activity (D'Auria, 2006; Tuominen et al., 2011; Bontpart et al., 2015). Based on protein-based phylogenetic analysis, the BAHD protein family is grouped to five clades, which are roughly differentiated by the type of substrates. Most Clade I members modify anthocyanin; Clade II is involved in the extension of long-chain epicuticular waxes; Clade III accepts a diverse range of alcohol substrates and preferred to use acyl-CoA as acyl donor; Clade V consists three subgroups involved in the biosynthesis of volatile ester, the compound paclitaxel, and lignin, respectively (D'Auria, 2006; Tuominen et al., 2011; Bontpart et al., 2015). By comparing transcriptional EST counts, Mitchell et al. (2007) found some BAHDs from Clade V, so called "Mitchell Clade", and glycosyltransferases (GTs) are highly expressed in grasses but barely expressed in dicots. This observation is in line with our assumption that enzymes for arabionxylan biosynthesis are supposed to be highly expressed in cereals and even substantially higher in grasses than in dicots to manifest reported structural differences in arabinoxylan of dicots and cereals cell wall (Mitchell et al., 2007).

Mitchell Clade of the BAHD acyltransferase (AT) gene family has 20 members in rice (Bartley et al., 2013) (**Figure 1.3**) and affects the formation of distinct ester conjugates by using acyl-CoA. One of these grass-diverged BAHD proteins, OsPMT (OsAT4 in our phylogenetic tree), was firstly supported by enzyme activity to be

responsible for the incorporation of pCA, rather than FA, into H and S monolignols (Withers et al., 2012). Later, OsAT4 was also introduced to *Arabidopsis* and poplar lignin, more likely creating G-*p*CA conjugates, which disagreed with its enzyme kinetic results in vitro, which showed S and H unit are preferable than G unit (Withers et al., 2012; Smith et al., 2015). Then, orthologous genes of *OsAT3* in *Brachypodium* (*BdPMT1*) were characterized genetically, showing that *BdPMT1* expression is positively related to cell wall lignin-*p*CA levels. And enhanced *BdPMT1* expression in *Brachypodium* has a greater impact on lignin polymerization by reducing lignin content and increasing S:G ratio, compared to knock down mutants (*PMT-RNAi* and *Bdpmt-1*), which do not significantly affect lignin composition, and improves enzyme digestibility of resulting plant biomass (Petrik et al., 2014). Heterologous expression of *BdPMT1* (*OsAT3*) and *BdPMT2* (*OsAT8*) in *Arabidopsis* secondary cell wall increased the amount of *p*CA-lignin, along with improvement in lignin alkaline solubility and biomass digestibility. BdPMT1/2-mediated acylation was found to occur mainly on sinapyl alcohol (S). Expression of *BdPMT1* in the *fah1* mutant deficient in S lignin synthesis accreted less *p*CA-lignin (*p*CA-G) than in wild type, additionally indicating that BdPMT1 is more likely to acylate *p*CA-CoA to S-lignin (Sibout et al., 2016). The generation of FA-monolignol in the *Arabidopsis ccrg1* mutant with *BdPMT1* expression also suggests the chance of using feruloyl-CoA as substrate (Sibout et al., 2016). AT3 in maize (*p*CAT) was identified by the purification of size-expected protein with sinapyl hydroxycinnamoyl activity (S-FA and S-*p*CA) (Marita et al., 2014). Then, the knock-down mutants also confirmed its p-coumaroyl monolignol activity and showed no effect on total lignin content but a decreased level of S units (Marita et al., 2014). Previous study

showed that overexpression of OsAT5 resulted in increased FA content in rice cell wall, suggesting that OsAT5 involved in incorporation of FA in cell wall polymers. In addition, OsAT5 shares the most identity with OsAT4. It becomes a strong candidate to decorate FA in lignin.

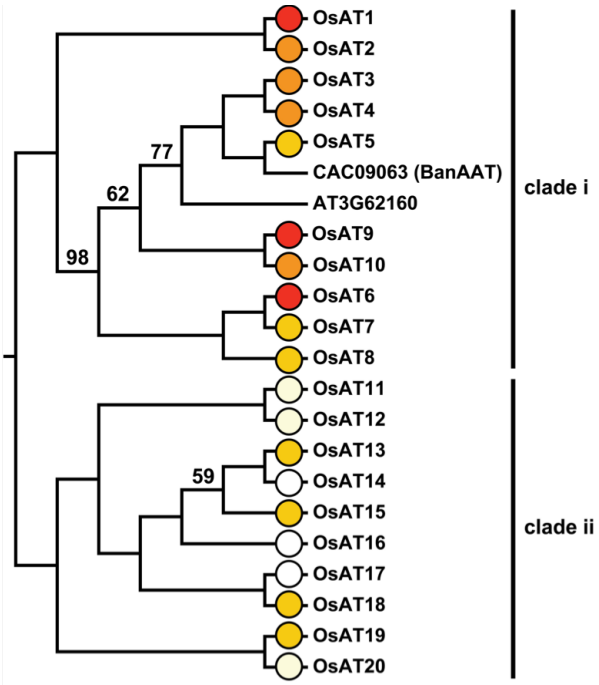


Figure 1.3 Simplified Mitchell Clade of the BAHD acyltransferase gene family in rice (Bartley et al., 2013).
The shading density of the circles on the tree branches indicates the level of RNA expression in terms of counts of Sanger ESTs and representation in massively parallel signature sequence data.

Apart from lignin acylation, some AT members appear to be distinctly involved in acylation of hemicellulose. Simultaneous knockdown of three BAHD members, *AT7*, *AT8*, *AT9*, and *AT10*, reduced ferulate in rice cell wall, suggesting that at least one of these ATs are involved in the synthesis of feruloyl arabinoxylan (Piston et al., 2010). Later on, Bartley et al. found that OsAT10 affects the bond formation between pCA and

arabinose in rice cell walls. Overexpression of *OsAT10* results in an increased arabinose-*p*CA and a reduction in polysaccharide-linked FA content, along with an improvement in biomass digestibility (Bartley et al., 2013). Overexpression of *BdAT1*, an ortholog gene of *OsAT1* in *Brachypodium*, also increased the feruloylation in cell wall (Buanafina et al., 2015). A recent study posited that *BdAT1* and *SvAT1*, which are orthologs of *OsAT9* in *Brachypodium* and *setaria* respectively, may be feruloyl arabinoxylan transferase (FAT) since their knockdown mutants significantly reduced feruloylation on arabinose and FA dimers (de Souza et al., 2018). However, the role of OsAT9 in rice cell wall decorations and rice cell wall recalcitrance is still unknown.

1.5 Aim and focus of the study

As aforementioned, lignocellulosic biofuel production is hindered by cell wall recalcitrance. Understanding characteristics and mechanisms of grass cell wall biosynthesis would be of great significance for the improvement of biomass utilization in agricultural and industrial settings. Hydroxycinnamates (*p*CA and FA), apart from featuring grass species, are considered as important factors of grass cell wall recalcitrance. Although previous studies have reported the role of some BAHD acyltransferase genes in *p*CA incorporation into grass cell walls; how cell wall polymers are feruloylated by FA and what BAHD acyltransferase genes are genetically responsible for that procedure are still open questions. Plus, we lack more knowledge of functional genes associated with cell wall features in less genetically tractable grass, especially the promising bioenergy crop switchgrass. My dissertation aimed to: 1) genetically profile the function of two rice BAHD acyltransferase genes, *OsAT5* and *OsAT9*; 2) examine the role of feruloyl lignin in biomass recalcitrance; 3) via transcriptomics, explore cell wall

recalcitrance genes in switchgrass. Major results of this study are presented in the following four chapters (2-5).

Chapter 2 presents analysis to the phylogeny of BAHD acyltransferase gene family and the prevalence of monolignol ferulates in a diversity of plant species. Feruloylation in plants will be reviewed from the perspective of evolution and enzymes responsible for that procedure will also be inferred.

Chapter 3 presents studies on *OsAT5* by heterologous expression of rice *OsAT5* in dicots and yeast, as well as chemical and enzymatic assays of its effect on cell wall structure, particularly feruloyl monolignol structure, and on cell wall recalcitrance in different organs and species.

Chapter 4 shows genetic characterization of *OsAT9* in rice by overexpression and knockdown, as well as analysis to its effect on the incorporation of FA and *p*CA into rice cell walls, xylan, and the digestibility of rice biomass.

Chapter 5 presents transcriptomic analysis of 48 switchgrass samples with varied digestibility and from different tissues, including leaf, upper internode, and lower internode. It covers the discovery of differentially expressed genes between genotypes and between tissues of interest, expression pattern analysis of cell wall biosynthesis genes, as well as gene ontology enrichment analysis.

1.6 References

Agbor VB, Cicek N, Sparling R, Berlin A, Levin DB (2011) Biomass pretreatment: Fundamentals toward application. Biotechnol Adv 29: 675–685

Akin DE (2008) Plant cell wall aromatics: influence on degradation of biomass. Biofuels, Bioprod Biorefining 2: 288–303

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, et al (2013) Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. Plant Physiol 161: 1615–1633

Besombes S, Mazeau K (2005a) The cellulose/lignin assembly assessed by molecular modeling. Part 1: Adsorption of a threo guaiacyl β-O-4 dimer onto a Iβ cellulose whisker. Plant Physiol Biochem 43: 299–308

Besombes S, Mazeau K (2005b) The cellulose/lignin assembly assessed by molecular modeling. Part 2: Seeking for evidence of organization of lignin molecules at the interface with cellulose. Plant Physiol Biochem 43: 277–286

Bevan MW, Garvin DF, Vogel JP (2010) Brachypodium distachyon genomics for sustainable food and fuel production. Curr Opin Biotechnol 21: 211–217

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54: 519–546

Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson N a, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M, et al (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. Nature 509: 376–80

Bontpart T, Cheynier V, Ageorges A, Terrier N (2015) BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. New Phytol 208: 695-707

Buanafina MM de O, Fescemyer HW, Sharma M, Shearer EA (2015) Functional testing of a PF02458 homologue of putative rice arabinoxylan feruloyl transferase genes in Brachypodium distachyon. Planta. 243: 659-674

Bunzel M, Heuermann B, Kim H, Ralph J (2008) Peroxidase-catalyzed oligomerization of ferulic acid esters. J Agric Food Chem 56: 10368–10375

Bunzel M, Ralph J, Lu F, Hatfield RD, Steinhart H (2004) Lignins and Ferulate−Coniferyl Alcohol Cross-Coupling Products in Cereal Grains. J Agric Food Chem 52: 6496–6502

Busse-Wicher M, Li A, Silveira RL, Pereira CS, Tryfona T, Gomes TCF, Skaf MS, Dupree P (2016) Evolution of xylan substitution patterns in gymnosperms and

angiosperms: implications for xylan interaction with cellulose. Plant Physiol 171: 2418-2431

Chávez Montes R a, Ranocha P, Martinez Y, Minic Z, Jouanin L, Marquis M, Saulnier L, Fulton LM, Cobbett CS, Bitton F, et al (2008) Cell wall modifications in Arabidopsis plants with altered alpha-L-arabinofuranosidase activity. Plant Physiol 147: 63–77

Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD, et al (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. Proc Natl Acad Sci U S A 109: 17117–22

Cosgrove DJ, Jarvis MC (2012) Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci. 3: 1-6

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9: 331–40

David K, Ragauskas AJ (2010) Switchgrass as an energy crop for biofuel production: A review of its ligno-cellulosic chemical properties. Energy Environ Sci 3: 1182–1190

Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. Plant Physiol 149: 137–41

Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P (2008) Land clearing and the biofuel carbon debt. Science (80-) 319: 1235–1238

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM (2006) Ethanol can contribute to energy and environmental goals. Science (80-) 311: 506–508

Grabber JH, Hatfield RD, Lu F, Ralph J (2008) Coniferyl ferulate incorporation into lignin enhances the alkaline delignification and enzymatic degradation of cell walls. Biomacromolecules 9: 2510–2516

Grabber JH, Hatfield RD, Ralph J (1998) Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. J Sci Food Agric 77: 193–200

Grantham NJ, Wurman-Rodrich J, Terrett OM, Lyczakowski JJ, Stott K, Iuga D, Simmons TJ, Durand-Tardif M, Brown SP, Dupree R, et al (2017) An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nat Plants 3: 859–865

Jiang B, Wang W, Gu F, Cao T, Jin Y (2016) Comparison of the substrate enzymatic digestibility and lignin structure of wheat straw stems and leaves pretreated by green liquor. Bioresour Technol 199: 181–187

Johar N, Ahmad I, Dufresne A (2012) Extraction, preparation and characterization of cellulose fibres and nanocrystals from rice husk. Ind Crops Prod 37: 93–99

Karlen SD, Free HCA, Padmakshan D, Smith BG, Ralph J, Harris PJ (2018) Commelinid Monocotyledon Lignins are Acylated by p-Coumarate. Plant Physiol. 177: 513-521

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan D, Helmich KE, Free HCA, Lee S, Smith BG, Lu F, et al (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci. Adv. 2:10

Krishnan A, Guiderdoni E, An G, Hsing Y -i. C, Han C -d., Lee MC, Yu S-M, Upadhyaya N, Ramachandran S, Zhang Q, et al (2009) Mutant Resources in Rice for Functional Genomics of the Grasses. Plant Physiol 149: 165–170

Lal R (2005) World crop residues production and implications of its use as a biofuel. Environ Int 31: 575–584

Lan W, Lu F, Regner M, Zhu Y, Rencoret J, Ralph SA, Zakai UI, Morreel K, Boerjan W, Ralph J (2015) Tricin, a Flavonoid Monomer in Monocot Lignification. Plant Physiol 167: 1284–1295

Li C, Knierim B, Manisseri C, Arora R, Scheller H V., Auer M, Vogel KP, Simmons BA, Singh S (2010a) Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. Bioresour Technol 101: 4900–4906

Li F, Zhang M, Guo K, Hu Z, Zhang R, Feng Y, Yi X, Zou W, Wang L, Wu C, et al (2015) High-level hemicellulosic arabinose predominately affects lignocellulose crystallinity for genetically enhancing both plant lodging resistance and biomass enzymatic digestibility in rice mutants. Plant Biotechnol J 13: 514–525

Li X, Ximenes E, Kim Y, Slininger M, Meilan R, Ladisch M, Chapple C (2010b) Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. Biotechnol Biofuels. 3:27

Lin F, Manisseri C, Fagerström A, Peck ML, Vega-Sánchez ME, Williams B, Chiniquy DM, Saha P, Pattathil S, Conlin B, et al (2016) Cell wall composition and candidate biosynthesis gene expression during rice development. Plant Cell Physiol 57: 2058–2075

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, et al (2008) How biotech can transform biofuels. Nat Biotechnol 26: 169–172

Mansfield SD, Kang K-Y, Chapple C (2012) Designed for deconstruction--poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. New Phytol 194: 91–101

Marita JM, Hatfield RD, Rancour DM, Frost KE (2014) Identification and suppression of the *p*-coumaroyl CoA:hydroxycinnamyl alcohol transferase in *Zea mays* L. Plant J 78: 850–864

Matias de Oliveira D, Finger-Teixeira A, Rodrigues Mota T, Salvador VH, Moreira-Vilar FC, Correa Molinari HB, Craig Mitchell RA, Marchiosi R, Ferrarese-Filho O, Dantas dos Santos W (2014) Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. Plant Biotechnol J., 13: 1224-1232

Mitchell R a C, Dupree P, Shewry PR (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. Plant Physiol 144: 43–53

de O Buanafina MM (2009) Feruloylation in grasses: current and future perspectives. Mol Plant 2: 861–72

Papa G, Varanasi P, Sun L, Cheng G, Stavila V, Holmes B, Simmons BA, Adani F, Singh S (2012) Exploring the effect of different plant lignin content and composition on ionic liquid pretreatment efficiency and enzymatic saccharification of Eucalyptus globulus L. mutants. Bioresour Technol 117: 352–359

Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. Biotechnol J 5: 147–162

Perlack RD, Wright LL, Turhollow AF, Graham RL, Forest BJS, Erbach DC (2005) BIOMASS AS FEEDSTOCK FOR A BIOENERGY AND BIOPRODUCTS INDUSTRY: THE TECHNICAL FEASIBILITY OF A BILLION-TON ANNUAL SUPPLY. U.S. Dep. Energy,Oak Ridge,TN

Petrik DL, Karlen SD, Cass CL, Padmakshan D, Lu F, Liu S, Le Bris P, Antelme S, Santoro N, Wilkerson CG, et al (2014) p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. Plant J 77: 713–26

Piston F, Uauy C, Fu L, Langston J, Labavitch J, Dubcovsky J (2010) Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. Planta 231: 677–91

Ragauskas AJ, Li M, Pu Y (2016) Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance. 4: 45

Ralph J (2010) Hydroxycinnamates in lignification. Phytochem Rev 9: 65–83

Del Río JC, Rencoret J, Prinsen P, Martínez ÁT, Ralph J, Gutiérrez A (2012) Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. J Agric Food Chem 60: 5922–5935

Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G (2006) Switchgrass as a biofuels feedstock in the USA. Can J Plant Sci 86: 1315–1325

Santos RB, Lee JM, Jameel H, Chang HM, Lucia LA (2012) Effects of hardwood structural and chemical characteristics on enzymatic hydrolysis for biofuel production.

Bioresour Technol 110: 232–238

Sarath G, Baird LM, Vogel KP, Mitchell RB (2007) Internode structure and cell wall composition in maturing tillers of switchgrass (Panicum virgatum. L). Bioresour Technol 98: 2985–2992

Saulnier L, Vigouroux J, Thibault JF (1995) Isolation and partial characterization of feruloylated oligosaccharides from maize bran. Carbohydr Res 272: 241–253

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61: 263–289

Schmer MR, Vogel KP, Mitchell RB, Perrin RK (2008) Net energy of cellulosic ethanol from switchgrass. Proc Natl Acad Sci 105: 464–469

Schmutz J, Schackwitz WS, Lipzen AM, Barry KW, Schmutz J (2016) Genome-Wide Sequencing of 41 Rice (Oryza sativa L.) Mutated Lines Reveals Diverse Mutations Induced by Fast-Neutron Irradiation. Mol Plant 9: 1078–1081

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F (2009) Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. Bioenergy Res 2: 233–245

Sibout R, Le Bris P, Legée F, Cézard L, Renault H, Lapierre C, Legee F, Cezard L, Renault H, Lapierre C, et al (2016) Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA:monolignol transferase (PMT). Plant Physiol 170: 1358-1366

Smith R a, Gonzales-Vigil E, Karlen SD, Park J-Y, Lu F, Wilkerson C, Samuels a. L, Ralph J, Mansfield SD (2015) Engineering monolignol p-coumarate conjugates into Poplar and Arabidopsis lignins. Plant Physiol 169: 2992-3001

Smith RA, Cass CL, Mazaheri M, Sekhon RS, Heckwolf M, Kaeppler H, de Leon N, Mansfield SD, Kaeppler SM, Sedbrook JC, et al (2017) Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol ferulates incorporated into maize lignins. Biotechnol Biofuels 10:109

Somerville C (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol 22: 53–78

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae); A worldwide phylogenetic classification of the Poaceae (Gramineae). J Syst Evol 53: 117–137

de Souza WR, Martins PK, Freeman J, Pellny TK, Michaelson L V., Sampaio BL, Vinecky F, Ribeiro AP, da Cunha BADB, Kobayashi AK, et al (2018) Suppression of a single BAHD gene in *Setaria viridis* causes large, stable decreases in cell wall feruloylation and increases biomass digestibility. New Phytol. 218: 81-93

Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE (2011) Lignin content in natural Populus variants affects sugar release. Proc Natl Acad Sci 108: 6300–6305

Takahama U, Oniki T (1994) Effects of ascorbate on the oxidation of derivatives of hydroxycinnamic acid and the mechanism of oxidation of sinapic acid by cell wall-bound peroxidases. Plant Cell Physiol 35: 593–600

Takeda Y, Koshiba T, Tobimatsu Y, Suzuki S, Murakami S, Yamamura M, Rahman MM, Takano T, Hattori T, Sakamoto M, et al (2017) Regulation of CONIFERALDEHYDE 5-HYDROXYLASE expression to modulate cell wall lignin structure in rice. Planta 246: 337–349

Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H (2003) Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol 133: 73–83

Terashima N, Awano T, Takabe K, Yoshida M (2004) Formation of macromolecular lignin in ginkgo xylem cell walls as observed by field emission scanning electron microscopy. Comptes Rendus - Biol 327: 903–910

Tu Y, Rochfort S, Liu Z, Ran Y, Griffith M, Badenhorst P, Louie G V, Bowman ME, Smith KF, Noel JP, et al (2010) Functional analyses of caffeic acid O-Methyltransferase and Cinnamoyl-CoA-reductase genes from perennial ryegrass (Lolium perenne). Plant Cell 22: 3357–73

Tuominen LK, Johnson VE, Tsai C-J (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. BMC Genomics 12: 236

U.S. DOE (2016) 2016 Billion-Ton Report. I: 411

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153: 895–905

Vogel KP, Dien BS, Jung HG, Casler MD, Masterson SD, Mitchell RB (2011) Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. Bioenergy Res 4: 96–110

Wang Y, Fan C, Hu H, Li Y, Sun D, Wang Y, Peng L (2016) Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops. Biotechnol Adv 34: 997–1017

Wende G, Fry SC (1997) 2-O-β-D-xylopyranosyl-(5-O-feruloyl)-L-arabinose, a widespread component of grass cell walls. Phytochemistry 44: 1019–1030

Wilkerson CG, Mansfield SD, Lu F, Withers S, Park J-Y, Karlen SD, Gonzales-Vigil E, Padmakshan D, Unda F, Rencoret J, et al (2014) Monolignol ferulate transferase

introduces chemically labile linkages into the lignin backbone. Science 344: 90–3

Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG (2012) Identification of Grass-specific Enzyme That Acylates Monolignols with p-Coumarate. J Biol Chem 287: 8347–8355

Wu Z, Zhang M, Wang L, Tu Y, Zhang J, Xie G, Zou W, Li F, Guo K, Li Q, et al (2013) Biomass digestibility is predominantly affected by three factors of wall polymer features distinctive in wheat accessions and rice mutants. Biotechnol Biofuels 6: 183

Zhang Y, Culhaoglu T, Pollet B, Melin C, Denoue D, Barrière Y, Baumberger S, Méchin V (2011) Impact of lignin structure and cell wall reticulation on maize cell wall degradability. J Agric Food Chem 59: 10129–10134

Chapter 2 Monolignol ferulate conjugates are naturally incorporated into plant lignins

Authors: Steven D. Karlen, **Chengcheng Zhang**, Matthew L. Peck, Rebecca A. Smith, Darshana Padmakshan, Kate E. Helmich, Heather C. A. Free, Seonghee Lee, Bronwen G. Smith, Fachuang Lu, John C. Sedbrook, Richard Sibout, John H. Grabber, Troy M. Runge, Kirankumar S. Mysore, Philip J. Harris, Laura E. Bartley and John Ralph

Publication Status: This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science Advances, Monolignol ferulate conjugates are naturally incorporated into plant lignins, Vol. 2, no. 10, e1600393, 14 Oct 2016, doi: 10.1126/scladv.1600393

Author contributions: CZ characterized *Ubi$_{pro}$::OsAT5* plants, including genotyping, gene expression, HCA analysis and lignin assays. CZ analyzed the DFRC data for both *OsAT5-D1* and *Ubi$_{pro}$::OsAT5* transgenic plants. CZ conducted phylogenetic analysis that found the convergent evolution of FMT. CZ contributed to writing this paper.

2.1 Abstract

Angiosperms represent most of the terrestrial plants and are the primary research focus for the conversion of biomass to liquid fuels and coproducts. Lignin limits our access to fibers and represents a large fraction of the chemical energy stored in plant cell walls. Recently, the incorporation of monolignol ferulates into lignin polymers was accomplished via the engineering of an exotic transferase into commercially relevant poplar. We report that various angiosperm species might have convergently evolved to natively produce lignins that incorporate monolignol ferulate conjugates. We show that this activity may be accomplished by a BAHD feruloyl-coenzyme A monolignol transferase, OsFMT1(AT5), in rice and its orthologs in other monocots.

2.2 Introduction

One of the major adaptations of terrestrial plants to life on land is their ability to produce lignin for structural strength and defense (Bonawitz and Chapple, 2010). Stochastically synthesized through stepwise radical coupling of 4-hydroxycinnamyl alcohols (called monolignols, primarily coniferyl and sinapyl alcohols), lignin is a polymer with aryl ether and various other C-O-C and C-C interunit connections (Boerjan et al., 2003; Ralph et al., 2004). Fragmentation of lignin allows polysaccharidase enzymes to access and convert cell wall polysaccharides to monomeric sugars or facilitates cell wall deconstruction to cellulose, hemicelluloses, and lignin fragments. These components are used in economically important processes, including production of paper and other fiber products, second-generation biofuels, and other bioproducts (FitzPatrick et al., 2010; Upton and Kasko, 2016). Lignin fragmentation often requires high temperatures and/or harsh chemical treatments to cleave even its weakest interunit bonds (Ragauskas et al., 2006; Somerville et al., 2010). However, if chemically labile ester bonds are introduced into the lignin polymer, as can be accomplished by augmenting the prototypical monomers with monolignol ferulate (ML-FA) conjugates (Figure 2.1A), then lignin fragmentation can occur under mild pretreatment conditions (Figure 2.1B) (Grabber et al., 2008; Ralph, 2009; Wilkerson et al., 2014). The findings here provide evidence that "zip-lignins," lignins derived, in part, from ML-FAs, have developed naturally via convergent evolution in diverse angiosperm lineages (Figure 2.2).

Initial work on the incorporation of ML-FAs into plant lignin was carried out by engineering poplar trees (*Populus alba* × *Populus grandidentata*) to express a gene from Chinese angelica [*Angelica sinensis* (*As*), a dicotyledonous Chinese medicinal plant]

encoding a feruloyl-coenzyme A (CoA) monolignol transferase (*As*FMT) (Wilkerson et al., 2014). The *As*FMT enzyme couples monolignols with feruloyl-CoA, an intermediate in the monolignol biosynthetic pathway (Figure S2.1), to produce ML-FAs. *As*FMT is a member of a family of proteins found in plants and fungi termed BAHD acyltransferases (AT; Figure 2.3) (D'Auria, 2006). *AsFMT*-expressing poplar trees produce ML-FAs and use them in lignification, resulting in improved cell wall saccharification following mild base pretreatment (Wilkerson et al., 2014).

The lignin assay, derivatization followed by reductive cleavage (DFRC) (Lu and Ralph, 1997), was used to confirm that ML-FAs were integrally incorporated into the lignin of *AsFMT*-expressing poplar trees (Wilkerson et al., 2014). DFRC cleaves b-ether bonds to release lignin fragments (Lu and Ralph, 1997), while leaving the ester linkages of incorporated ML-FAs intact. Hence, DFRC-releasable coniferyl and sinapyl dihydroferulate diacetates (ML-DHFAs; Figure 2.1C) are diagnostic for the ML-FAs incorporated into lignin and their release level is proportional to the amount of ML-FAs in the lignin (Wilkerson et al., 2014). However, because DFRC releases just a fraction of the incorporated ML-FAs as ML-DHFAs, only the relative level of conjugates in the lignin can be determined (Wilkerson et al., 2014). The threshold to detect DFRC-released ML-DHFAs by gas chromatography-multiple reaction monitoring-mass spectrometry (GC-MRM-MS; Table S2.1) is ~0.01 mg/g of acetyl bromide soluble lignin (ABSL) (see Materials and Methods). These experimental conditions revealed that wild-type (WT) poplar trees already release low levels (0.3 mg/g of ABSL) of the ML-DHFAs from their lignins (Wilkerson et al., 2014), indicating that poplar plants naturally synthesize ML-FA conjugates and use them in lignification.

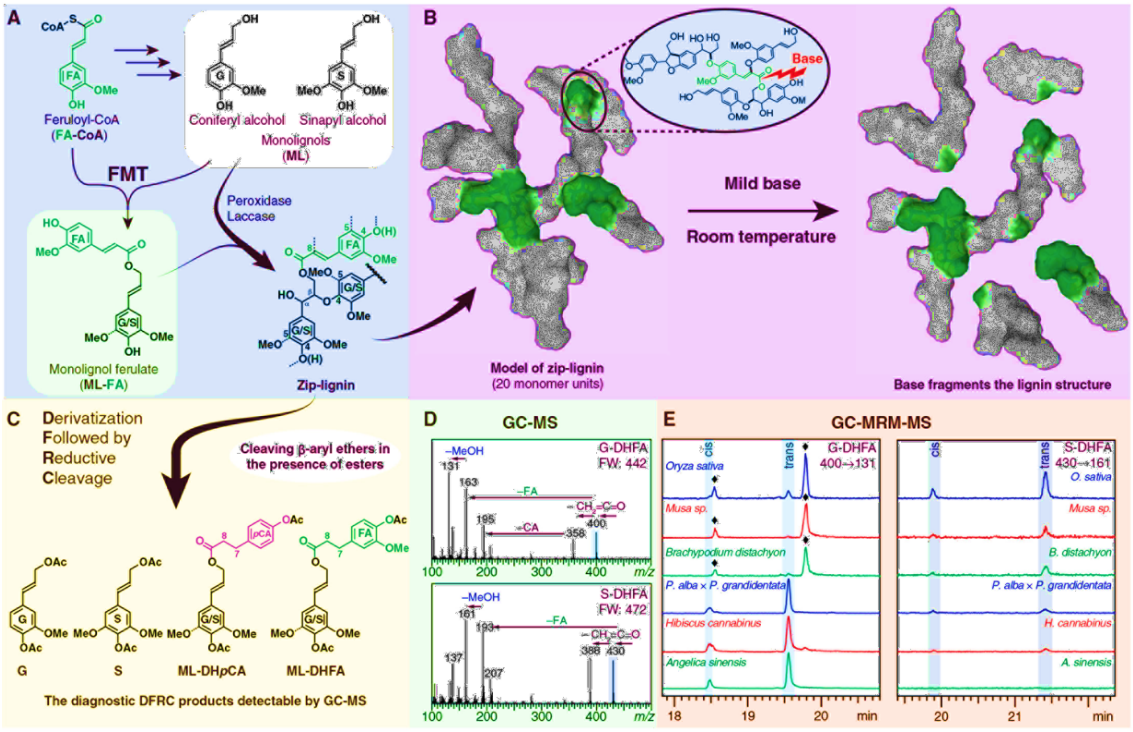


Figure 2.1 Incorporation of ML-FAs into lignin introduces chemically labile esters into the polymer backbone.

(A) FMT enzyme couples feruloyl-CoA and monolignols together to form ML-FA conjugates. The compounds are then transported to the cell wall and undergo radical coupling-based polymerization to form lignin; all the bonds that can be formed when ML-FAs are incorporated into b-ether structures in zip-lignin are shown with dashed lines. **(B)** Mild base (for example, 0.05 M NaOH at 30°C) cleaves the ML-FA-derived (green) ester bonds dividing the polymer into ≤ (n + 1) fragments, where n is the number of ML-FA units. **(C)** DFRC breaks down the lignin by cleaving b-aryl ethers but leaving the esters intact. **(D)** Electron impact MS fragmentation pattern for coniferyl and sinapyl DHFA (G-DHFA and S-DHFA). FW, formula weight; m/z, mass/charge ratio. **(E)** GC-MRM-MS chromatograms of the DFRC product mix reveal the presence of the diagnostic products for ML-FA incorporation into lignin from a number of WT plants. The symbol ♦ indicates the signals corresponding to S-DH*p*CA, which shares an MRM transition with G-DHFA.

2.3 Results

2.3.1 Plants accumulating ML-FAs in extractives also use them for lignification

Armed with the new knowledge that plants naturally incorporate ML-FAs and an analytical method to diagnostically detect their incorporation into the polymer with suitable sensitivity, we reexamined the lignin of three plants known to produce ML-FAs in their extractives: Chinese angelica (*A. sinensis*, family Apiaceae), kenaf (*Hibiscus cannabinus*, family Malvaceae), and balsa (*Ochroma pyramidale*, family Malvaceae). Performed on isolated lignins, the DFRC assay showed that these plants also used ML-FAs in their lignification (Table S2.2).

2.3.2 Phylogeny of plants using ML-FAs in lignification

The abovementioned discovery prompted a survey to determine whether the utilization of ML-FAs in lignification was a trait of eudicots or all angiosperms or whether it was a trait ubiquitous to all lignifying plants. We performed the DFRC assay on a set of plants representing the spermatophytes or "seed plants," including 13 gymnosperms and 54 angiosperms (Figure 2.2A). The assayed gymnosperms showed no evidence of ML-FAs, whereas low levels were present in the lignin of many, but not all, of the angiosperms (Figure 2.2B and Tables S2.3 to S2.5). Some of the assayed eudicots showed detectable levels of the diagnostic ML-DHFAs, predominantly derived from coniferyl ferulate (G-FA). Many of the monocots, especially the recently evolved commelinids, which include the major cereal crop plants, tested positive for ML-FAs, with sinapyl ferulate (S-FA) as the main conjugate.

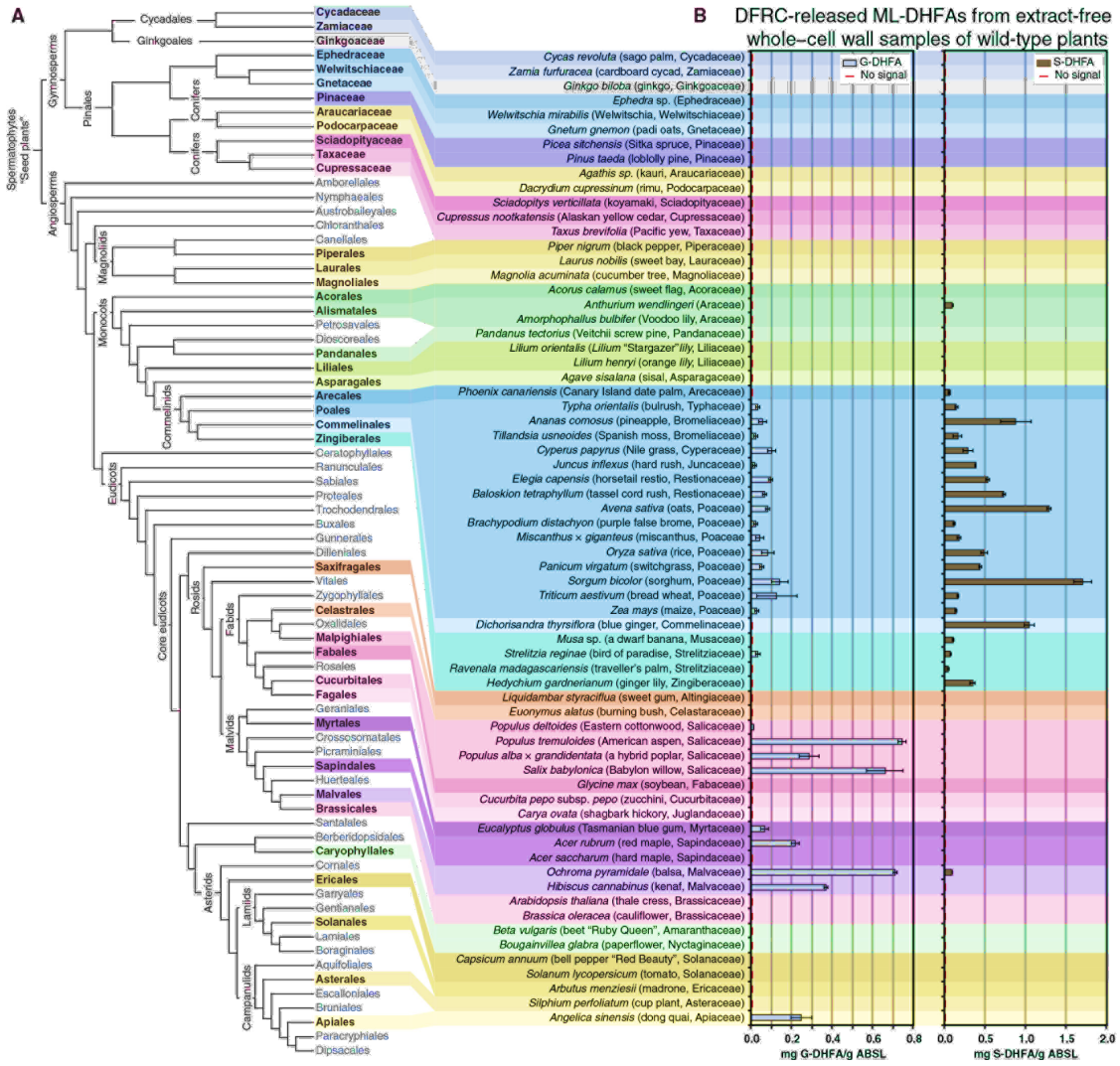


Figure 2.2 Comparison of the DFRC-releasable ML-DHFA conjugates among plant species.
(A) A phylogenetic tree of the spermatophytes ("seed plants"), with the orders and families in which plant species were studied. **(B)** DFRC-released ML-DHFA conjugates; red bars indicate no evidence of ML-DHFAs. Bars indicate SEM for the summation of detected conjugates on duplicate analyses run on a single sample prepared from each plant species.

2.3.3 Origin of ML-FAs and transferase specificity

The discovery of ML-FA production in this diversity of plant species, and notably in plants that use other types of monolignol conjugates for lignification, prompted us to investigate whether ML-FAs arise from a lack of specificity in other transferases or from

dedicated FMT enzymes. The "nonspecific transferases" hypothesis is most readily addressed in grasses, namely, in rice [*Oryza sativa* (Os)] and *Brachypodium* [*Brachypodium distachyon* (Bd)], in which *p*-coumaroyl-CoA monolignol transferases (PMTs), as well as BAHD ATs, have already been identified [*AT4* = *OsPMT1* (Withers et al., 2012), Bradi2g36910 = *BdPMT1* (Petrik et al., 2014; Withers et al., 2012); and Bradi2g36980 = *BdPMT2* (Petrik et al., 2014; Sibout et al., 2016; Withers et al., 2012).

To test the specificity of PMT, we compared the amount of DFRC-releasable ML-DHFAs between a *Brachypodium* line with native BdPMT1 expression and a sodium azide-generated Bdpmt missense mutant line that produced lignins devoid of p-coumarates (Petrik et al., 2014). Levels of DFRC-released ML-DHFA were similar between the *Brachypodium* WT (0.14 mg/g of ABSL) and mutant lines (0.15 mg/g of ABSL) (Figure 2.4 and Table S2.6), indicating that *Bd*PMT1 activity does not influence ML-FA production and, thus, ML-FAs cannot be attributed to the low specificity of *Bd*PMT1.

Furthermore, introduction of the *OsPMT1* gene from rice into the eudicot *Arabidopsis* (*Arabidopsis thaliana*), which does not natively incorporate detectable levels of monolignol conjugates of any kind into its lignin (<0.01 mg/g of ABSL), resulted in transformants with *p*-coumarates esterified to lignin at a level that was easily quantified by DFRC through the release of monolignol dihydro-*p*-coumarates (ML-DH*p*CAs) (Smith et al., 2015). However, both the *Os*PMT1-expressing and WT *Arabidopsis* lines did not produce any detectable ML-DHFAs (Figure 2.4 and Table S2.6). These two experiments strongly suggest that ML-FA conjugates incorporated into commelinid monocot lignins are not the result of low PMT specificity.

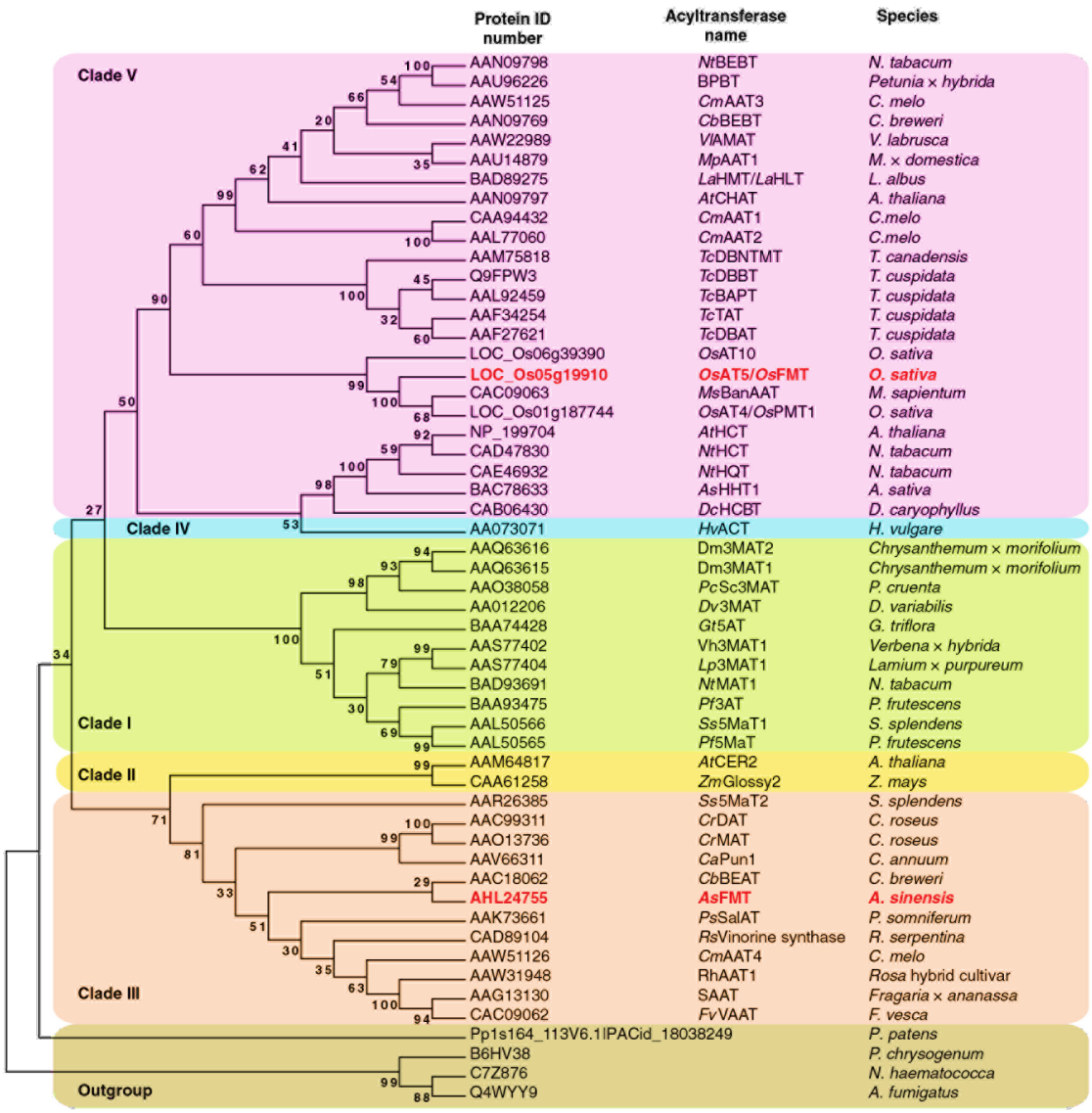


Figure 2.3 Phylogenic reconstruction of BAHD acyl-CoA ATs.
It is consistent with the convergent evolution of the two feruloyl-CoA monolignol transferases, *Os*AT5/FMT and *As*FMT. Maximum likelihood phylogeny of *As*FMT, *Os*AT5, and biochemically characterized BAHD proteins (D'Auria, 2006). Branch values are based on 1000 bootstraps. Protein IDs are National Center for Biotechnology Information GenBank identifiers or genome locus identifiers. Species codes for locus identifiers with order classifications in parentheses are as follows: AT, *Arabidopsis thaliana* (Brassicales); Bradi, *Brachypodium distachyon* (Poales); Eucgr, *Eucalyptus grandis* (Myrtales); Glyma, *Glycine max* (Fabales); GSMUA, *Musa acuminata* (Zingiberales); GRMZM, *Zea mays* (Poales); LOC_Os, *O. sativa* (Poales); Medtr, *Medicago truncatula* (Fabales); Pavir, *Panicum virgatum* (Poales); PDK, *Phoenix dactylifera* (Arecales); POTR, *Populus trichocarpa* (Malpighiales); and Sb, *Sorghum bicolor* (Poales).

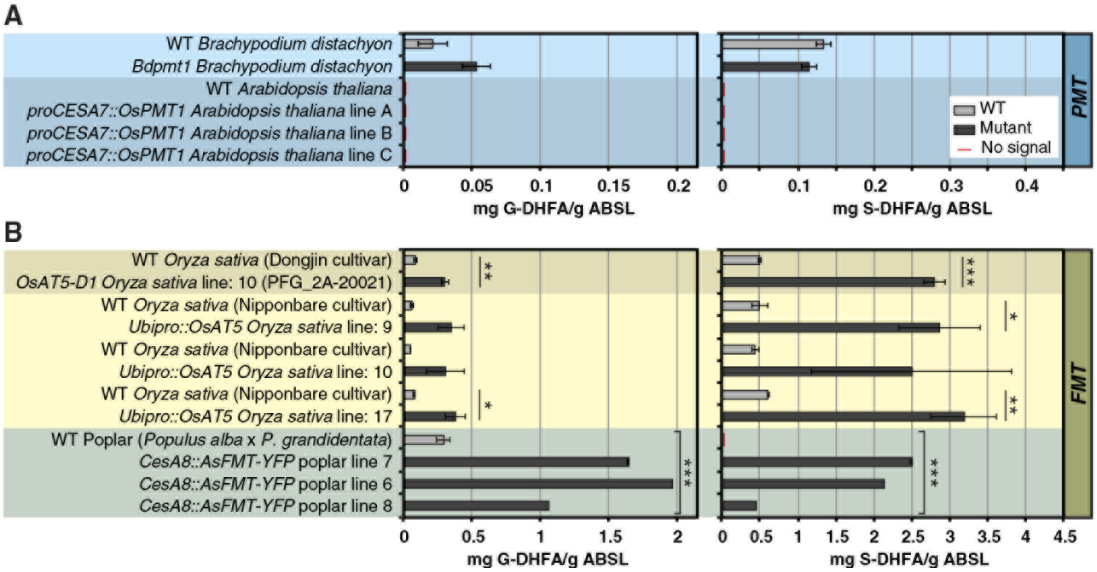


Figure 2.4 The amounts of DFRC-releasable ML-DHFA conjugates correlate with the expression of *FMT* genes but not with the expression of *PMT*.
(A) No significant change was observed between WT Brachypodium and a *Bdpmt* mutant with no PMT activity. Introduction of *BdPMT* into *Arabidopsis* results in detectable ML-DH*p*CA but no detectable MLDHFA. **(B)** Rice overexpressing *OsAT5* (*OsFMT1*), either via activation-tagging in *OsAT5-D1* or via a *Ubi* promoter, and transgenic *AsFMT* poplar show an increase (five- to seven-fold) in ML-DHFAs. Bars indicate SEM of three to seven biological replicates that were measured with technical replicates for each. *P < 0.05, **0.001 < P < 0.01, and ***P < 0.001, Student's t test.

2.3.4 Identification of a putative new FMT in commelinids

To test the hypothesis that ML-FAs are the product of dedicated FMT enzymes, we examined the annotated genomes of members of the grass family Poaceae, in which ML-FAs appear to be ubiquitous, for proteins with close homology to *As*FMT and *Os*PMT1.We find that these proteins are in distinct BAHD subclades (Figure 2.3, clades III and V, respectively). Although rice and other commelinids lack homologs of *As*FMT (Figure S2.2A), they have multiple genes encoding proteins with similarity to PMT (Figure S2.2B) (Bartley et al., 2013). From screening of rice BAHD mutant lines, we previously reported the preliminary observation that an activation tagged rice mutant

genotype, *OsAT5-D1*, has increased ferulic acid in cell walls compared to WT individuals (Bartley et al., 2013). *Os*AT5 is among the most similar proteins to *Os*PMT1 (56% identity) and thus seemed to be a reasonable candidate for an enzyme having FMT activity.

Here, we provide molecular genetic evidence consistent with the hypothesis that *Os*AT5 is an FMT. The *OsAT5-D1* line has a transfer DNA (T-DNA) insertion with multiple copies of a 35S transcription enhancer upstream of the *OsAT5* transcriptional start site (Figure S2.3A). We confirmed the increased expression of *OsAT5* in rice plants with the T-DNA insertion compared with negative segregants lacking the insertion (Figure S2.3B). In contrast, the expression of more distant genes flanking the insertion site remained unchanged (Figure S2.3). As hypothesized, *OsAT5-D1* plants exhibited an approximately fivefold increase in the DFRC-releasable ML-DHFA but no increase in ML-DH*p*CA; by comparison, the best *AsFMT* poplar lines showed an approximately sevenfold increase in released ML-DHFAs (Figure 2.4B and Table S2.6) (Wilkerson et al., 2014). The increase in ferulate is not associated with arabinoxylans, a well-known location of ferulate incorporation in grass cell walls (Harris and Trethewey, 2009), because ferulate released by mild acid hydrolysis remained unchanged (Figure S2.5). To confirm the association between increased expression of *Os*AT5 and increased lignin-associated ferulate esters, we generated and characterized three additional overexpression lines, in which *OsAT5* expression was driven by the *ZmUbi* promoter (Figure 2.4B and Figure S2.4). All showed increased DFRC-releasable ML-DHFA (Figure 2.4B and Table S2.6), compared to the negative segregant controls, establishing that OsAT5 acts on the monolignols. By all indications, then, *Os*AT5 is a native FMT enzyme in rice, that is, *Os*AT5 = *Os*FMT1.

Having identified *OsAT5* as a potential grass FMT gene, we compared the *OsAT5* sequence to known BAHD AT gene sequences in other commelinid monocots. BAHD proteins were previously identified in the annotations of several plant species (Bartley et al., 2013), and here, we add the data from the draft genomes of banana and date palm (both nongrass commelinids), maize, switchgrass, and eucalyptus. All commelinids examined have several *Os*AT5/FMT homologs, whereas this subgroup of ATs has only one or two homologs in the eudicots (Figure S2.2B). The presence of *OsAT5*-encoding genes corresponds with detectable levels of DFRC-releasable ML-DHFAs, suggesting that the commelinid monocot FMT genes may have evolved as a trait of this group.

2.4 Discussion

Upon examination of the phylogeny of the assessed plant species (Figure 2.2), some patterns emerged. For example, plants known to produce significant levels of other monolignol conjugates also incorporate ML-FAs into their lignins. These include the commelinids, which incorporate monolignol *p*-coumarates into their lignins and also have ferulates acylating the arabinoxylan hemicelluloses (Harris and Trethewey, 2009), and plants that use monolignol p-hydroxybenzoates in lignification, such as palms (*Arecaceae*), willows (*Salix*), and poplars/aspen (*Populus*). A weaker link exists between plants incorporating ML-FAs and those known to use monolignol acetates, which include kenaf and sisal (Del Río et al., 2007), several other monocots (Rencoret et al., 2013; del Río et al., 2012), and various hardwoods (Ralph and Lu, 1998). However, the introduction of ML-FAs into *Arabidopsis* via transgenesis, which appears to natively lack them, indicates that the ability to use monolignol conjugates does not appear to be isolated to

those species that have evolved to do so, highlighting the plasticity of cell wall lignification.

The evidence for BAHD proteins with FMT activity (Figure S2.2) suggests that the activity has arisen at least twice; that is, it has convergently evolved. For example, the two nonhomologous *FMT* protein sequences, *Os*FMT from a commelinid monocot and *As*FMT from a eudicot, have little sequence similarity (20%). Furthermore, the model that shows that FMT activity is an ancestral trait of angiosperms is inconsistent with the apparent absence of DFRC-releasable ML-DHFAs in most of the non-commelinid monocots and many of the eudicots, although it cannot be ruled out. Whether an *As*FMT ortholog is responsible for the incorporation of ML-FAs into eudicot lignins is unclear from the phylogenetic reconstructions. Analysis of the genomes of the ML-FA-producing eudicot tree species *Eucalyptus globulus* and *Populus trichocarpa* reveals a large number of clade III BAHD ATs that are absent from the commelinids, none of which have great similarity to *As*FMT, as the closest poplar homologs are POPTR_0001s31750 and POPTR_0004s01720, with 36% identity (Figure S2.2A).

The convergent evolution and subsequent proliferation of plants that incorporate ML-FA conjugates into their lignins indicate that, potentially, there is a biological advantage for the production of this lignin structure. Regardless of the actual driving forces selecting for them, the diversity and environmental success of plants with native zip-lignins show that they have no apparent general disadvantages in terms of plant defense or structural stability.

Finally, our findings further refute the contention by some researchers that lignins are derived only from three monolignols. It has been increasingly evident over the past

20 years that many other compounds biosynthesized by plants are used as monomers in lignification (Boerjan et al., 2003; Mottiar et al., 2016; Ralph et al., 2004; Simmons et al., 2010). As demonstrated here, ML-FA conjugates must now be added to the list of authentic lignin precursors. In practical terms, our discovery unveils new approaches to increasing levels of readily cleavable ester bonds in the lignin backbone, either by breeding or by transgenic methods similar to those used to introduce *AsFMT* into poplar (Wilkerson et al., 2014). Further work is also needed to explore the effects of ML-FA-containing lignins on processes such as carbon sequestration and biomass utilization.

2.5 Materials and Methods

2.5.1 DFRC procedure for ML-DHFA conjugates released from cell wall lignins

Incorporation of ML-FAs into the lignin was determined using the ether-cleaving, ester-retaining DFRC method previously established for ML-*p*CA and ML-FA conjugates (Lu, 2014; Lu and Ralph, 1999; Petrik et al., 2014; Wilkerson et al., 2014). Extract-free cell wall samples (50 mg) or enzyme lignin samples (50 mg) were stirred in 2-dram vials fitted with polytetrafluoroethylene pressure release caps with acetyl bromide/acetic acid [1:4 (v/v), 4 ml]. After heating for 3 hours at 50°C, the solvents were removed by SpeedVac (50°C, 35 min, 1.0 torr, 35 torr/min; Thermo Scientific SPD131DDA). Crude films were suspended in absolute ethanol (0.5 ml), dried on SpeedVac (50°C, 15 min, 6.0 torr, 35 torr/min), and then suspended in dioxane/acetic acid/water [5:4:1 (v/v), 5 ml] with nanopowder zinc (250 mg). The vials were then sealed, sonicated to ensure suspension of solids, and stirred in the dark at room temperature for 16 to 20 hours. Additional nanopowder zinc was added as required to maintain a fine suspension of zinc in the reaction mixtures. The reaction mixtures were then quantitatively transferred with

dichloromethane (DCM; 6 ml) into separatory funnels charged with a saturated ammonium chloride (10 ml) and an internal standard (diethyl 5,5′-diferulate diacetate, 54.0 mg). Organics were extracted with DCM (4 × 10 ml), combined, dried over anhydrous sodium sulfate, and filtered, and the solvents were removed via rotary evaporation (water bath at <50°C). Free hydroxyl groups on DFRC products were then acetylated for 16 hours in the dark using a solution of pyridine and acetic anhydride [1:1 (v/v), 5 ml], after which the solvents were removed via a rotary evaporator to yield crude oily films.

To remove most of the polysaccharide-derived products, acetylated DFRC products were loaded onto solid phase extraction (SPE) cartridges (Supelco Supelclean LC-Si SPE tube, 3 ml, product no. 505048) with DCM (2 × 1.0 ml). After elution with hexanes/ethyl acetate [1:1 (v:v), 8 ml], solvents were removed by rotary evaporation, and the products were transferred in stages with GC-MS–grade DCM to final sample volumes of 200 ul into GC-MS vials containing a 300-ul insert. Samples were analyzed on a triple-quadrupole GC-MS/MS (Shimadzu GCMS-TQ8030) operating in MRM mode using synthetic standards for authentication and calibration. The GC program and acquisition parameters are listed in Table S2.2.1, and the result of the DFRC assay is listed in tables S2 to S4.

2.5.2 Detection threshold for DFRC-released ML-DHFAs

When DFRC products are analyzed via a triple-quadrupole GC-MS, operating in MRM mode, the detectable amount of releasable ML-DHFA is ~0.01 mg/g of ABSL. This estimated threshold of detection is partially based on instrument limitations but is mainly limited by the presence of a large amount of polysaccharides, lignin fragments, and the

DFRC-released monolignols that create a complex matrix from which the desired ML-DHFA products need to be extracted. Scaling up the DFRC reaction increases not only the amount of ML-DHFA products formed but also the amount of the other matrix components. We have found that the crossover point for improved sensitivity with increasing biomass is ~50 to 100 mg of plant cell walls.

2.5.3 Estimating the ML-FA incorporation levels

Most ML-FAs incorporated into lignin form structures that are not cleaved by DFRC and are not released as measurable ML-DHFAs products. Therefore, biomimetically lignified cell walls prepared with known amounts of ML-FAs were subjected to DFRC to estimate the efficiency of ML-DHFA production from ML-FAs incorporated into plant lignins. For this purpose, primary maize cell walls containing bound native peroxidases were lignified by the dropwise addition of separate solutions of dilute hydrogen peroxide and a 1:1 mixture of coniferyl and sinapyl alcohols substituted with either G-FA [0, 8.4, 15.4, and 26.7 weight % (wt%)] or S-FA (0, 9.0, 16.5, and 28.4 wt%) (10). The artificially lignified cell walls were then analyzed by DFRC, and the results were plotted as the weight % of ML-FA used to prepare the cell walls versus the weight % of ML-DHFA released per gram of ABSL. Regression constants for the released ML-DHFAs were determined by fitting the data to a linear model ($y = mx$, where y = wt% ML-FA in cell wall dehydrogenation polymer and x = wt% ML-DHFA released; G-FA: m = 53, R^2 =0.99; S-FA: m = 34, R^2 = 0.99). These results indicate a direct relationship between DFRC-releasable ML-DH conjugates and the quantity of ML-FA in lignin. However, we stress that this approach provides only a rough means of estimating the quantity of ML-FAs incorporated into native plant lignins.

Applying these results to the yields of ML-DHFAs from surveyed plant species suggests that ML-FA incorporation into lignin averaged 1.8 wt% and ranged from 0.2 to 6.6 wt% (tables S2 to S4). The average weight % of ML-FA in eudicots (2.0 wt%) was slightly higher than the average for the monocots (1.7 wt%). The monocots on average released more S-DHFA than G-DHFA (77:23) as compared to the eudicots that mostly released G-DHFA (S-DHFA/G-DHFA, 1:99).

2.5.4 Procedure to determine ABSL content

The ABSL contents were measured in 1-cm quartz cuvettes on a Shimadzu UV-1800 spectrophotometer at l = 280 nm and e280 = 20.0, as previously described (Fukushima and Hatfield, 2001; Hatfield et al., 1999).

2.5.5 Preparation of cell wall and enzyme lignin samples

Cell wall samples were prepared by one of four methods (A to D) described below.

Method A: Gymnosperm and angiosperm samples were freeze-dried and shaker-milled into fine powder (Retsch MM 400). Samples were then solvent-extracted sequentially with water (3 × 45 ml), 80% ethanol (3 × 45 ml), and acetone (2 × 45 ml) by first suspending the sample in solvent, sonicating for 20 min, pelleting by centrifugation (8800g for 20 min; Sorvall Biofuge Primo Centrifuge), and, finally, decanting the supernatant. The extract-free pellet was then dried under vacuum for analysis.

Method B: Tissues with lignified walls were identified histochemically [phloroglucinol-HCl and ultraviolet fluorescence microscopy (Smith and Harris, 1995)]. The lignified tissues were isolated and ground (ROCKLABS Ltd.) in 3-(N-morpholino) propanesulfonic acid-KOH buffer [20 mM (pH6.8)]. The homogenates were centrifuged (1000g for 10 min); resuspended in buffer; filtered through a nylon mesh (11-um pore

size); washed successively with buffer, ethanol, methanol, and n-pentane; and air-dried. Starch granules were detected and removed from the preparations from *Cyperus papyrus, Hedychium gardnerianum,* and *Phoenix canariensis*, using the method of Carnachan and Harris (Carnachan and Harris, 2000).

Method C: Hardwood eudicot samples were Wiley-milled to pass through a 40-mesh screen, air-dried, and extracted three times with a 9:1 mixture of acetone and water. Three additional extractions with methanol were used when required to remove intensely colored extractives. After air-drying, the samples were again Wiley-milled to pass through a 40-mesh screen.

Method D: Samples of mature stems were processed to destarched alcohol-insoluble residues (AIR), following a reported procedure (Bartley et al., 2013).

Enzyme lignins were prepared from ball-milled materials, as previously described (Chang et al., 1975; Wagner et al., 2007). Briefly, ball-milled extract-free materials (1 g; prepared using method A) in 50-ml centrifuge tubes were incubated at 35°C and mixed in a shaker at 225 rpm for 3 days with 40 ml of sodium acetate buffer (pH 5.0) and 40 mg of crude cellulases (CELLULYSIN, EMD Biosciences). After incubation, solids were pelleted by centrifugation (8800g for 20 min; Sorvall Biofuge Primo Centrifuge) and washed with acetate buffer (2 × 40 ml). Washed solids were then treated again with the crude cellulases (40 mg) for 3 days and then pelleted and washed with reverse osmosis water (3 × 40 ml). The resulting pelleted solids were dried on a freeze dryer to yield ~10% of the dry weight of the original ball-milled cell wall material.

2.5.6 Phylogenetic analysis of BAHD ATs

Phylogenetic analysis was conducted in two stages. First, to determine the evolutionary relationship between *Os*AT5 and *As*FMT proteins, we conducted a maximum likelihood phylogenetic reconstruction of three characterized rice ATs (*Os*AT5, *Os*AT4/PMT, and *Os*AT10) and *As*FMT with a set of 46 biochemically characterized BAHD enzymes that had previously been divided into five clades, I to V (D'Auria, 2006), and three PF02458-containing proteins from fungi and one from moss (*Physcomitrella patens*) to serve as an outgroup. We used MEGA5.2.2 (Tamura et al., 2011) to infer and visualize maximum likelihood phylogenies with the following parameters: amino acid substitutions according to the Jones-Taylor-Thornton model, g distribution of mutation rate among sites, a distribution shape parameter of 5, and gaps treated by partial deletion, allowing site coverage as low as 95%. From these analyses (Figure 2.3), we concluded that *As*FMT and *Os*AT5 are not evolutionarily related. Rather, *Os*AT5, *Os*AT4, and *Os*AT10 are closely related to the *Ms*BanAAT in BAHD clade V and belong to a subclade of BAHD proteins, previously referred to as the "Mitchell clade." By contrast, *As*FMT is a member of clade III.

Second, we separately analyzed clade III and the Mitchell clade across diverse species and focused on the genera represented in the DFRC screening (Figure S2.2, A and B). We included two nongrass commelinid monocots, banana and palm, because of the expectation that these plants might have a close homolog of the FMT in grasses. To identify putative BAHD ATs from the diverse species, we used HMMER version 3.1 (Finn et al., 2011) with the hidden Markov model profile for PF02458 from the Pfam database. We searched the following genome annotation sources and versions, which

were current at the time of the analysis: maize (*Zea mays*), MaizeGDB version 2; switchgrass (*Panicum virgatum*), Phytozome version 1.1; palm (*Phoenix dactylifera*), Weill Cornell Medicine-Qatar PDK sequence version 3; eucalyptus (*Eucalyptus grandis*), Phytozome version 1.1; and banana (*Musa acuminata*), Banana Genome Hub DH-Pahang v1. For comparison with previous analyses, we also included sequences from the following sources: *Arabidopsis* (*A. thaliana*), The Arabidopsis Information Resource version 10; soybean (*Glycine max*), Phytozome version 7.0; *Medicago truncatula*, Mt3.5; sorghum (*Sorghum bicolor*), Phytozome version 7.0; rice (*O. sativa*), Michigan State University version 6.1; and *B. distachyon*, Phytozome version 7.0 (Bartley et al., 2013). Phylogenetic analyses were then conducted with the Pfam domain protein sequences (Figure S2.2B). We determined the BAHD clade of each predicted protein via comparison with the D'Auria set using Clustal2 (Larkin et al., 2007) and omitted sequences that lacked the region surrounding the highly conserved active-site motif HXXXD. From the maximum likelihood phylogenetic reconstructions for these initial single species (100 bootstraps), we identified proteins most closely related to those in the Mitchell clade and clade III. For clade III, we identified 12 proteins from eucalyptus, 9 from *Arabidopsis*,1from palm, and none from rice, maize, or banana, suggesting that clade III is not as widely represented in commelinids as in dicots (Figure S2.2A). For the Mitchell clade, we identified 22 predicted proteins from banana, 20 from rice, 17 from maize, 23 from switchgrass, 13 from sorghum, 16 from *Brachypodium*, 11 from palm, and 1 from eucalyptus. This builds on the previous results by providing evidence that this clade of genes not only is more abundant in grasses thanineudicotyledonous plants (Bartley et al., 2013) but also includes numerous members in other commelinid monocots.

The four subgroups (a to d) of Mitchell clade "i" (within clade V), which have both grass and banana and/or palm proteins, are consistent with there being at least four members of this gene family in the last common ancestor between the grasses and other commelinids. We then carried out maximum likelihood phylogenetic reconstructions with the commelinid expanded Mitchell clade and related protein sequences and the dicot-expanded clade III. We used 1000 bootstraps for the Mitchell clade tree, with the outgroup consisting of the *Arabidopsis* spermidine dicoumaroyl transferase and spermidine disinapoyl transferase proteins and a group of closely related enzymes that function in taxol biosynthesis. For the clade III tree, we ran 500 bootstraps, and the outgroup consisted of two clade II proteins.

2.5.7 Gene expression of OsPMT *in A. thaliana*

The *OsPMT* gene from rice was introduced into *Arabidopsis* (*A. thaliana*), which does not natively incorporate monolignol conjugates of any kind into its lignin, and targeted to secondary cell wall-forming cells (*proCELLUOSE SYNTHASE 7::OsPMT*) (Smith et al., 2015). These plant lines successfully produced monolignol *p*-coumarates and incorporated them into their lignins. DNA sequences are shown in Text S2.1.

2.5.8 Rice lines overexpressing OsFMT (OsAT5)

Rice plants were grown in a greenhouse in Turface Athletics medium/ vermiculite (1:1) mix supplemented three times per week with fertilizer (Jack's Professional LX 15-5-15 4Ca 2Mg) at temperatures from 29° to 32°C during the day and from 24° to 25°C during the night. After germination in water, 7-day-old seedlings were transplanted to the greenhouse. Natural day lengths of less than 13 hours were supplemented with artificial lighting.

We characterized two classes of *OsFMT/OsAT5* overexpression rice mutants, an activation-tagged line and three independent Ubiquitin1 promoter lines. Depending on the assay, we characterized the second (T2) and/or third (T3) generation of selfed progeny of activation-tagged homozygous mutant and WT segregant rice line of *PFG_2A-20021*, referred to herein as *OsAT5-D1*.The T1 generation of this line, which is in the Dongjin cultivar background, was previously reported to have an increase in wall-associated ferulates in the mature leaf sheaths (Bartley et al., 2013). We characterized mutant progeny of *2A-20021.10* and WT segregant progeny of *2A-20021.11*. DFRC data are from theT3 generation (mutant 2A20021.10.7.45.# and WT 2A-20021.11.2.40.#). Trifluoroacetic acid (TFA) fractionation data are from the T2 generation (mutant 2A20021.10.2.# and WT 2A-20021.11.7.#), as are gene expression data. In this notation, the period separates generation identities, and the numbers indicate the number of the parental plant of the analyzed progeny. Genotyping, as previously described by Bartley et al. (Bartley et al., 2013), was used to isolate homozygous lines and for spot-checking to confirm future generations.

To generate independent lines overexpressing *OsAT5*, a pUC-based plasmid containing the full coding sequence for *OsAT5* was synthesized de novo. We then recombined the gene into the *pCAMBIA1300-Ubi-GW-Nos* construct (Park et al., 2010) to produce *pCAMBIA1300-UbiAT5*. This binary vector contains a Gateway cassette, flanked by the maize *Ubi1* promoter, the 3′-terminator of nopaline synthase from *Agrobacterium tumefaciens*, and the *Hpt2* gene that confers resistance to hygromycin. The rice cultivar Nippon-bare was used for transformation after introducing the overexpression vector *pCAMBIA1300-UbiAT5* into the *A. tumefaciens* strain EH105. The

transformation procedure was similar to the method described by Nishimura et al. (Nishimura et al., 2006). Briefly, embryogenic calli were obtained from 3-week-old immature embryos of Nipponbare. EH105 suspended at an optical density of 0.1 at 600 nm (approximately 3×10^6 colony-forming units/ml) was cocultivated with calli for 90 s and placed on sterile filter paper to remove excess Agrobacterium. The infected calli were incubated for 60 hrs in the dark and, later, were washed and reincubated on regeneration media for 4 weeks under continuous light (3.5 klux). Regenerated plants were grown in deep plastic containers for 1 week, transferred to pots for growth in the greenhouse under the conditions described above, and genotyped with primers for *Hpt2* and the *Ubi$_{pro}$::AT5* construct (Table S2.7). We characterized the T1 progeny of *Ubi$_{pr}$o::AT5* lines-9, 10, and 17-along with negative segregant WT plants for each line. Several additional lines were set aside because of low T0 seed set or apparent lack of expression of the transgene, assayed as described below.

OsAT5 overexpression was assayed using quantitative real-time reverse transcription polymerase chain reaction (Figure S2.3 and S2.4), with SYBR green and primers as previously described (Table S2.7) (Bartley et al., 2013; Jain, 2009; Piston et al., 2010). Two reference genes, *Ubq5* and *Cc55*, were used for every sample measured. Gene expression was measured in leaf tissue from vegetatively developing rice plants at approximately the V5 stage. We selected samples from the same leaf and developmental stage, that is, developmentally matched, between controls and mutants for each line. For the *OsAT5D1* line, samples were T2 mutant and nontransgenic negative segregants approximately 5 weeks after sowing. For the *Ubi$_{pro}$::AT5* lines, samples were second-

generation transgenic (T1) and negative segregants lacking the transgene from mutant lines 9, 10, and 17.

2.5.9 Analysis of carbohydrate-associated ferulic acid of OsAT5-D1

We used weak acid fractionation to determine whether the changes in hydroxycinnamate content were associated with matrix polysaccharides or lignin. Destarched AIR was created as previously described (Bartley et al., 2013; Saulnier et al., 1995). For weak acid treatment, 2.5 mg of destarched AIR was mixed with 500 ul of either 0.05 M TFA or water, similar to a previously described method (Bartley et al., 2013). Samples were incubated with shaking at 100°C for varying times. At each time point, the supernatant (containing solubilized cell wall material) was separated from the remaining solid by centrifugation, and materials were frozen to stop the reaction. Thawed samples were treated with 2 M NaOH and neutralized with concentrated HCl, trans-cinnamic acid was added as an internal standard, and the samples were then ethyl acetate–extracted and analyzed by high-performance liquid chromatography as previously described (Bartley et al., 2013).

2.6 References

Bartley, L.E., Peck, M.L., Kim, S.-R., Ebert, B., Manisseri, C., Chiniquy, D.M., Sykes, R., Gao, L., Rautengarten, C., Vega-Sánchez, M.E., et al. (2013). Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. Plant Physiol. *161*, 1615–1633.

Boerjan, W., Ralph, J., and Baucher, M. (2003). LIGNINBIOSYNTHESIS. Annu. Rev. Plant Biol. *54*, 519–546.

Bonawitz, N.D., and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. Annu. Rev. Genet. *44*, 337–363.

Carnachan, S.M., and Harris, P.J. (2000). Polysaccharide compositions of primary cell walls of the palms Phoenix canariensis and Rhopalostylis sapida. Plant Physiol. Biochem. *38*, 699–708.

Chang, H.-M., Cowling, E.B., and Brown, W. (1975). Comparative Studies on Cellulolytic Enzyme Lignin and Milled Wood Lignin of Sweetgum and Spruce. Holzforschung *29*, 153–159.

D'Auria, J.C. (2006). Acyltransferases in plants: a good time to be BAHD. Curr. Opin. Plant Biol. *9*, 331–340.

Del Río, J.C., Marques, G., Rencoret, J., Martínez, A.T., and Gutiérrez, A. (2007). Occurrence of naturally acetylated lignin units. J. Agric. Food Chem. *55*, 5461–5468.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

FitzPatrick, M., Champagne, P., Cunningham, M.F., and Whitney, R.A. (2010). A biorefinery processing perspective: treatment of lignocellulosic materials for the production of value-added products. Bioresour. Technol. *101*, 8915–8922.

Fukushima, R.S., and Hatfield, R.D. (2001). Extraction and isolation of lignin for utilization as a standard to determine lignin concentration using the acetyl bromide spectrophotometric method. J. Agric. Food Chem. *49*, 3133–3139.

Grabber, J.H., Hatfield, R.D., Lu, F., and Ralph, J. (2008). Coniferyl ferulate incorporation into lignin enhances the alkaline delignification and enzymatic degradation of cell walls. Biomacromolecules *9*, 2510–2516.

Harris, P.J., and Trethewey, J.A.K. (2009). The distribution of ester-linked ferulic acid in the cell walls of angiosperms. Phytochem. Rev. *9*, 19–33.

Hatfield, R.D., Grabber, J., Ralph, J., and Brei, K. (1999). Using the acetyl bromide assay to determine lignin concentrations in herbaceous plants: some cautionary notes. J. Agric. Food Chem. *47*, 628–632.

Jain, M. (2009). Genome-wide identification of novel internal control genes for normalization of gene expression during various stages of development in rice. Plant Sci. *176*, 702–706.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lu, F. (2014). Lignin: Structural Analysis, Applications in Biomaterials and Ecological Significance (Nova Science Pub Incorporated).

Lu, F., and Ralph, J. (1997). DFRC Method for Lignin Analysis. 1. New Method for β-Aryl Ether Cleavage: Lignin Model Studies. J. Agric. Food Chem. *45*, 4655–4660.

Lu, F., and Ralph, J. (1999). Detection and determination of p-coumaroylated units in lignins. J. Agric. Food Chem. *47*, 1988–1992.

Mottiar, Y., Vanholme, R., Boerjan, W., Ralph, J., and Mansfield, S.D. (2016). Designer lignins: harnessing the plasticity of lignification. Curr. Opin. Biotechnol. *37*, 190–200.

Nishimura, A., Aichi, I., and Matsuoka, M. (2006). A protocol for Agrobacterium-mediated transformation in rice. Nat. Protoc. *1*, 2796–2802.

Park, C.-J., Bart, R., Chern, M., Canlas, P.E., Bai, W., and Ronald, P.C. (2010). Overexpression of the endoplasmic reticulum chaperone BiP3 regulates XA21-mediated innate immunity in rice. PLoS One *5*, e9262.

Petrik, D.L., Karlen, S.D., Cass, C.L., Padmakshan, D., Lu, F., Liu, S., Le Bris, P., Antelme, S., Santoro, N., Wilkerson, C.G., et al. (2014). p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. Plant J. *77*, 713–726.

Piston, F., Uauy, C., Fu, L., Langston, J., Labavitch, J., and Dubcovsky, J. (2010). Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. Planta *231*, 677–691.

Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr, Hallett, J.P., Leak, D.J., Liotta, C.L., et al. (2006). The path forward for biofuels and biomaterials. Science *311*, 484–489.

Ralph, J. (2009). Hydroxycinnamates in lignification. Phytochem. Rev. *9*, 65–83.

Ralph, J., and Lu, F. (1998). The DFRC Method for Lignin Analysis. 6. A Simple Modification for Identifying Natural Acetates on Lignins. J. Agric. Food Chem. *46*, 4616–4619.

Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P.F., Marita, J.M., Hatfield, R.D., Ralph, S.A., Christensen, J.H., et al. (2004). Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. Phytochem. Rev. *3*, 29–60.

Rencoret, J., Ralph, J., Marques, G., Gutiérrez, A., Martínez, Á.T., and del Río, J.C. (2013). Structural characterization of lignin isolated from coconut (Cocos nucifera) coir fibers. J. Agric. Food Chem. *61*, 2434–2445.

del Río, J.C., Prinsen, P., Rencoret, J., Nieto, L., Jiménez-Barbero, J., Ralph, J., Martínez, A.T., and Gutiérrez, A. (2012). Structural characterization of the lignin in the cortex and pith of elephant grass (Pennisetum purpureum) stems. J. Agric. Food Chem. *60*, 3619–3634.

Saulnier, L., Vigouroux, J., and Thibault, J.F. (1995). Isolation and partial characterization of feruloylated oligosaccharides from maize bran. Carbohydr. Res. *272*, 241–253.

Sibout, R., Le Bris, P., Legée, F., Cézard, L., Renault, H., and Lapierre, C. (2016). Structural Redesigning Arabidopsis Lignins into Alkali-Soluble Lignins through the Expression of p-Coumaroyl-CoA:Monolignol Transferase PMT. Plant Physiol. *170*, 1358–1366.

Simmons, B.A., Loqué, D., and Ralph, J. (2010). Advances in modifying lignin for enhanced biofuel production. Curr. Opin. Plant Biol. *13*, 313–320.

Smith, B.G., and Harris, P.J. (1995). Polysaccharide composition of unlignified cell walls of pineapple [Ananas comosus (L.) Merr.] fruit. Plant Physiol. *107*, 1399–1409.

Smith, R.A., Gonzales-Vigil, E., Karlen, S.D., Park, J.-Y., Lu, F., Wilkerson, C.G., Samuels, L., Ralph, J., and Mansfield, S.D. (2015). Engineering Monolignol p-Coumarate Conjugates into Poplar and Arabidopsis Lignins. Plant Physiol. *169*, 2992–3001.

Somerville, C., Youngs, H., Taylor, C., Davis, S.C., and Long, S.P. (2010). Feedstocks for lignocellulosic biofuels. Science *329*, 790–792.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

Upton, B.M., and Kasko, A.M. (2016). Strategies for the Conversion of Lignin to High-Value Polymeric Materials: Review and Perspective. Chem. Rev. *116*, 2275–2306.

Wagner, A., Ralph, J., Akiyama, T., Flint, H., Phillips, L., Torr, K., Nanayakkara, B., and Te Kiri, L. (2007). Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in Pinus radiata. Proc. Natl. Acad. Sci. U. S. A. *104*, 11856–11861.

Wilkerson, C.G., Mansfield, S.D., Lu, F., Withers, S., Park, J.-Y., Karlen, S.D., Gonzales-Vigil, E., Padmakshan, D., Unda, F., Rencoret, J., et al. (2014). Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. Science *344*, 90–93.

Withers, S., Lu, F., Kim, H., Zhu, Y., Ralph, J., and Wilkerson, C.G. (2012). Identification of grass-specific enzyme that acylates monolignols with p-coumarate. J. Biol. Chem. *287*, 8347–8355.

Chapter 3 Expression of a rice ferulate monolignol transferase in *Arabidopsis* improves cell wall suitability for biorefining

Authors: Chengcheng Zhang, Alex Tsai, Rebecca A. Smith, Steven D. Karlen, Matthew L. Peck, Mary F. LaPorte, Nick Santoro, John Ralph, Henrik Scheller, Laura E. Bartley

Publication Status: This chapter is in preparation as a manuscript for publication

Author Contributions: This was a collaborative project with Joint BioEnergy Institute and Department of Energy Great Lakes Bioenergy Research Center. CZ coordinated closely with collaborators and designed experiments. CZ performed the research on transgenic *Arabidopsis* plants, including plasmid constructs, generation of transgenic plants, characterization of T1 generation, and cell wall analysis. CZ analyzed most of the data and wrote the paper.

3.1 Abstract

Lignin in cell walls limits bioconversion of plant biomass to biofuels and bioproducts. Engineering plant lignin promises to reduce cell wall recalcitrance. Our previous study identified a rice "BAHD" acyltransferase, OsAT5, whose overexpression increased the abundance of feruloyl monolignol esters (ML-FAs) in lignin. Here we confirmed the function of OsAT5 in producing ML-FAs in *Saccharomyces cerevisiae* and *Arabidopsis*, both of which are not known to naturally form these esters. When incorporated into lignin polymers, ML-FA conjugates might render cell wall structures less recalcitrant to mild base pretreatment, due to base-cleavable ester bonds. However, *AT5-D1* rice straw did not exhibit reduced cell wall recalcitrance, whereas *Arabidopsis* $C4H_{pro}$*-OsAT5* showed a 23-41% reduction in enzymatic deconstruction with alkaline pretreatment but not with hot water pretreatment. We also observed other differences between these transgenics, depending on the organ examined. Sinapyl ferulate (S-FA) and coniferyl ferulate (G-FA) both increased in rice straw 5-fold while in Ubi_{pro}*-OsAT5* rice roots, the quantitative increase in S-FA was similar, but the increase in G-FA was 10-fold. In contrast, transgenic *Arabidopsis* incorporated more S-FA than G-FA in the inflorescence stem lignin. $C4H_{pro}$*-OsAT5 Arabidopsis* lignin was reduced slightly by 14%, but also incorporated less S lignin (a 30% decrease, $p=0.003$), though similar changes were not observed in the rice overexpression lines. These discordant observations in rice and *Arabidopsis* indicate that ML-FAs produced by OsAT5 have differential impacts depending on plant species and tissues. Despite the increased ML-FAs in both transgenic rice and *Arabidopsis*, wild-type rice lignin with more ML-FA conjugates presented a 2-fold higher alkaline solubility than wild-type *Arabidopsis* lignin, indicative of distinct

polymer arrangements and/or size in rice lignin. This study shows the potential use of OsAT5 in structurally modifying plant cell walls and producing less recalcitrant plant biomass for biorefining. The differential recalcitrance in grasses and dicots as a result of increasing in ML-FA conjugates also raises the possibility of plant/lignin structure-specific engineering.

3.2 Introduction

Lignocellulosic biomass represents a sustainable resource for biofuel and biochemical production. Technologies for biomass-to-biofuels conversion have been developed and advanced upon a century (Ragauskas et al., 2006). However, the challenge remains to efficiently convert cellulose of plant cell walls to cost-competitive biofuels due to cell wall recalcitrance, which is the natural resistance of the complex cell wall matrix to biological deconstruction (Himmel et al., 2007). Plant cell walls mainly contain a mixture of sugar-based polysaccharides (35-50% cellulose and 25-30% hemicellulose) and lignin (15-30%) (Ragauskas et al., 2006). Lignin embeds within the cell wall and interacts with other biopolymers, and thus represents a major contributor to recalcitrance (Boerjan et al., 2003; Vanholme et al., 2010). The lignin polymer is composed of phenolics and appears to have evolved with the adaptation of plants to a terrestrial life for mechanical support and UV protection (Vanholme et al., 2010). The main building blocks of lignin include coniferyl alcohol, sinapyl alcohol, and *p*-coumaryl alcohol; they differ in the degree of methoxylation on phenol rings and generate guaicyl (G), syringyl (S), and *p*-hydroxyphenyl (H) after incorporated to lignin polymers. Reducing lignin content, lignin hydrophobicity, and its bonding with itself and sugar polysaccharides, can improve enzymatic hydrolysis of plant cell walls (Ragauskas et al., 2016).

Ferulates (FA) and *para*-coumarates (*p*CA) are two major hydroxycinnamates in grass lignin. As intermediates of the lignin biosynthetic pathway, these phenolic acids are also attached to cell wall polymers and may alter biomass recalcitrance. *p*CA is mostly found as a terminal pendent on lignin, possibly serving as “radical” catalysts to facilitate the polymerization of S monomers (Ralph, 2010). The enzymes responsible for

esterification of *p*CA in lignin has been characterized both biochemically and genetically (Grabber et al., 1998; Withers et al., 2012; Marita et al., 2014; Petrik et al., 2014; Smith et al., 2015; Sibout et al., 2016). Increased *p*CA -monolignol conjugates in Brachypodium lignin improves biomass digestibility (Petrik et al., 2014). Introducing *p*CA into *Arabidopsis* lignin also improves the alkaline solubility of lignin and enzymatic hydrolysis without pretreatment (Sibout et al., 2016).

FA esters in grasses, mainly in the form of feruloyl arabinoxylan, can also involve in lignin polymerization. FA was firstly proposed to act as nucleation sites to initiate lignification (Bunzel et al., 2004; Ralph, 2010). The resulting crosslinking between arabinoxylan and lignin is one of the most important factors conferring grass cell wall recalcitrance (Grabber et al., 1998). FA-ester conjugates are recently considered as a new lignin building block, FA moiety of which can participate in radical coupling reactions during lignin polymerization and in turn facilitate lignin fragmentation under mild base pretreatment, so called "zip-lignin". This concept was firstly hypothesized in a maize cell wall model study showing that the incorporation of coniferyl ferulate into lignin enhanced alkaline-based delignification and enzymatic degradation (Grabber et al., 2008; Ralph, 2010). The zip-lignin model was confirmed by introducing a feruloyl-monolignol transferase (FMT) from Chinese angelica [*Angelica sinensis* (*As*), a dicotyledonous Chinese medicinal plant] into polar. Heterologous expression of *As*FMT driven by *CesA8* promoter in poplar tree (*Populus alba* × *Populus grandidentata*) produces monolignol ferulates (ML-FAs), which are then incorporated into lignin polymers. The resulting biomass presented improved saccharification after mild base pretreatment (Wilkerson et al., 2014). Similarly, suppression of the first lignin specific biosynthetic enzyme,

cinnamoyl-CoA reductase, increased the intercellular pool of feruloyl-CoA and ML-FAs in maize lignin polymers, but decreased lignin content and lignin monomers along with the enhanced digestibility of stem rind tissue (Smith et al., 2017a). ML-FAs are also found in natural lignin of many terrestrial plants, including some dicots and most of commelinid monocot (e.g., Arecales, Poales, Commelinales, and Zingiberales). Among the dicots producing these conjugates, most of them only have coniferyl ferulates (G-FA), with balsa (*Ochroma pyramidale*) as an exception, which also contains a small amount of sinapyl ferulates (S-FA). In contrast, all commelinid monocots examined exclusively contain S-FA at a level higher than G-FA (Karlen et al., 2016). We reported that overexpression of *OsAT5*/*OsFMT*, a BAHD acyltransferase in rice increased both G-FA and S-FA conjugates in lignin polymers of mature rice straw, suggesting that OsAT5 may be a feruloyl monolignol transferase. Similar studies also found that constitutive overexpression of *OsAT5* in rice increased both S-FA and G-FA with the same extent in aboveground tissues (Karlen et al., 2016). Poplar does not have an ortholog gene of *OsAT5* but endogenously produces G-FA in lignin, indicating that plants may have convergently evolved the ability to synthesize this conjugate with unknown genes (Karlen et al., 2016).

Here, we investigated the biochemical function of OsAT5 and dissected the impact of OsAT5 and ML-FAs on lignin synthesis and biomass digestibility. First, we provide strong support for OsAT5 functioning as an FMT through heterologous expression in yeast and the dicot, *Arabidopsis thaliana*. We then examined cell wall properties of transgenic rice and *Arabidopsis,* including biomass digestibility, lignin content and composition, as well as lignin solubility. Interestingly, we found increasing

ML-FA conjugates in plant lignin elicits a discordant effect on the preference of using ML-FAs and the digestibility of transgenic biomass. The results increase our understanding of lignin biosynthesis and suggest new plant/lignin-specific engineering strategies to improve biomass utilization.

3.3 Results

3.3.1 The evidence of feruloyl-monolignol transferase activity of OsAT5 *in yeast*

Previously, we showed that overexpression of *OsAT5*/*OsFMT* in rice (*OsAT5-D1* and three Ubi_{pro}-*OsAT5* lines) increased both G-FA and S-FA conjugates in lignin polymers of mature rice straw, suggesting that OsAT5 may be a feruloyl monolignol transferase. To biochemically characterize its enzymatic activity, we expressed *OsAT5* in a yeast (*Saccharomyces cerevisiae*) co-expression system, which demonstrate to generate monolignol hydroxycinnamates via co-expression of the BAHD acyltransferase and *Arabidopsis* 4-coumaryl: CoA ligase (At4CL5) (Eudes et al., 2016b). In this genetically engineered system, yeast is fed with donor and acceptor molecules. At4CL5 can then activate the donor through esterification of coenzyme A. If it possesses the hypothesized enzymatic activity, OsAT5 will transfer the coenzyme A activated donor onto the acceptor (**Figure 3.1A**). We obtained authenticated coniferyl ferulate as our standard of liquid chromatography-mass spectrometry (LC-MS) measurement, and readouts from mass spectrometry ionization fragmentation, retention time, and *m/z* ratios, are used to infer the formation of acyltransferase products. Coniferyl ferulate standard was fragmentized to generate ferulic acid fragment, which presented the same retention time (4.308 min) as the authenticated conjugated standard (**Figure 3.1B and C**). Presumably, permutations of hydroxycinnamic acid and hydroxycinnamoyl alcohol would result in a

similar spectral structure as coniferyl ferulate, but with a lower degree of methoxylation. Such chemical changes would also reduce the hydrophobicity of resulting conjugates and shorten the retention time compared to coniferyl ferulate.

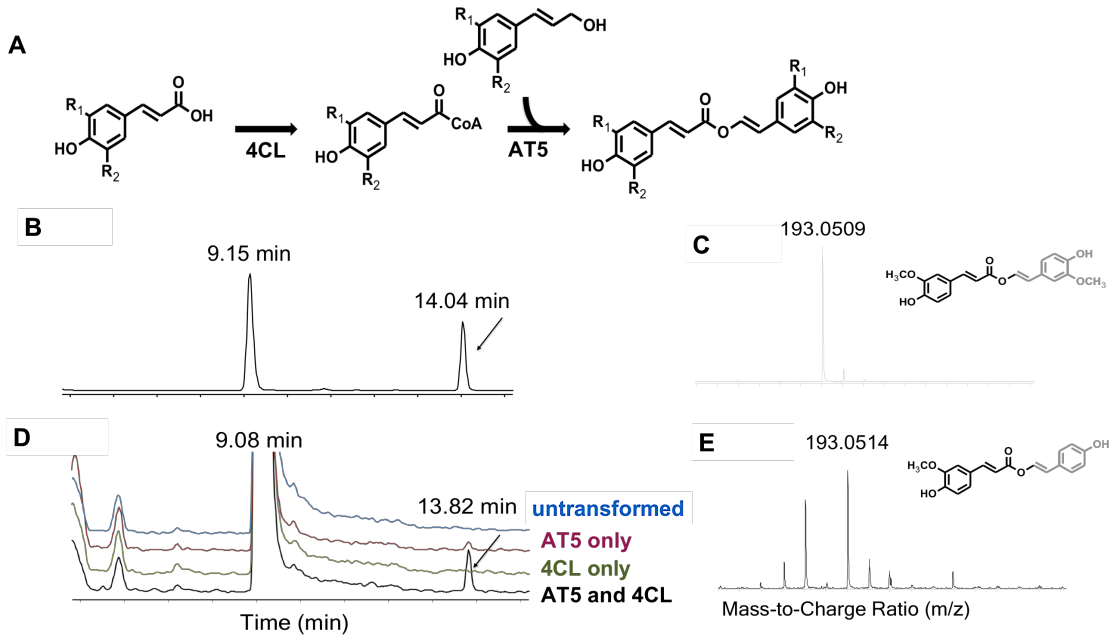


Figure 3.1 Feeding monolignols to yeast that co-expresses OsAT5 and At4CL5 generates the coniferyl ferulate conjugate.
(A) A scheme of enzyme reactions by 4CL and OsAT5. **(B)** LC chromatograms and **(C)** MS spectra of coniferyl ferulate standard. **(D)** LC chromatograms and **(E)** MS spectra of *pDRf1-4CL5-OsAT5* yeast culture with negative controls including untransformed yeast culture, *pDRf1-OsAT5* yeast culture, and *pDRf1-At4CL5* yeast culture.

As a control of the yeast-based study, we ran a test on a characterized *p*-coumaroyl CoA: monolignol transferase, OsAT4, as our positive control for OsAT5 profiling, which showed a similar high substrate binding affinity for *p*-coumaryl alcohol (H) and sinapyl alcohol (S) in in vitro enzyme assays (Withers et al., 2012; Smith et al., 2015). Yeast that constitutively co-express both OsAT4 and At4CL5 were constructed and cultured in presence of *p*-coumaric acid and *p*-coumaryl alcohol. MS analysis of

intracellular extracts demonstrated the synthesis of *p*-coumaryl coumarate (**Supplemental Figure S3.1a**). This is consistent with the previous OsAT4 in vitro analysis that *p*-coumaryl alcohol and *p*CA-CoA have the highest affinity (Withers et al., 2012). This product was absent when only At4CL5 was expressed (**Supplemental Figure S3.2a**), suggesting it was not formed by endogenous enzymes in yeast or by At4CL5 alone. Similarly, production of coniferyl coumarate was observed when yeast was fed with *p*-coumaric acid and coniferyl alcohol (**Supplemental Figure S3.1b**); a weak production of coniferyl ferulate was also detected when ferulic acid and coniferyl alcohol were present simultaneously (**Supplemental Figure S3.1d**). Both *p*-coumaric coumarate and *p*-coumaric ferulate are identifiable by MS (**Supplemental Figure S3.1**), with retention times at 4.174 min and 4.277 min respectively. Surprisingly, we did not observe the synthesis of sinapyl coumarate when feeding yeast with *p*-coumaric acid and sinapyl alcohol, even though OsAT4 had a high binding affinity to these substrates in in vitro assays (Withers et al., 2012).

With this set-up, engineered yeast co-expressing OsAT5 and At4CL5 was cultured in presence of ferulic acid and coniferyl alcohol. As expected, coniferyl ferulate was exclusively synthesized by OsAT5 and identified by MS (**Figure 3.1D and E**) but not detected in the control, which only expressed At4CL5 (**Figure 3.1D, Supplemental Figure S3.2**). In contrast to OsAT4, OsAT5 did not show detectable acyltransferase activity when fed with *p*-coumaric acid as the acyl donor (**Supplemental Figure S3.3a, b**). We also did not see sinapyl ferulate when sinapyl alcohol was provided as the acyl acceptor (**Supplemental Figure S3.3d**). It is clear that OsAT5 exhibited a substrate profile distinct from that of OsAT4. OsAT5 only forms hydroxycinnamoyl conjugates

upon feeding with ferulic acid and coniferyl alcohol, which defines its feruloyl-CoA: monolignol acyltransferase activity.

3.3.2 Ubi_{pro}-OsAT5 *lines increased lignin feruloylation in rice root*

Since cell wall composition and structure vary with cell types and tissues, using straw materials that include both leaf and stem tissues may confound possible tissue-specific effects on lignin. To more specifically study the role of *OsAT5* in grass lignification, we examined *OsAT5* overexpression-caused changes in the cell wall of root where *OsAT5* has a relatively high indigenous expression (**Figure 3.2A**). Whole roots from three *Ubi_{pro}-OsAT5* rice lines and corresponding wild types at the post-reproductive stage were sampled for conjugates measurement using DFRC (Derivative followed by reductive cleavage), which can cleave the lignin signature β-ether bonds while leave γ-esters intact. Compared with the level of G-FA (0.5 mg/g) and S-FA (13 mg/g) in the wildtype root, *Ubi_{pro}-OsAT5* root presented a 10-fold increase in G-FA (4.7 mg/g) and a 4-fold increase in S-FA (49 mg/g) (**Figure 3.2B**). The change in G-FA is even more obvious in root than in straw (a ~5-fold increase) we previously reported in *Ubi_{pro}-OsAT5* lines (Karlen et al., 2016). However, the abundance of *p*-coumarate monolignol conjugates was not changed there (**Supplementary Figure S3.4)**. These data provide additional evidence supporting that AT5 is a feruloyl monolignol transferase.

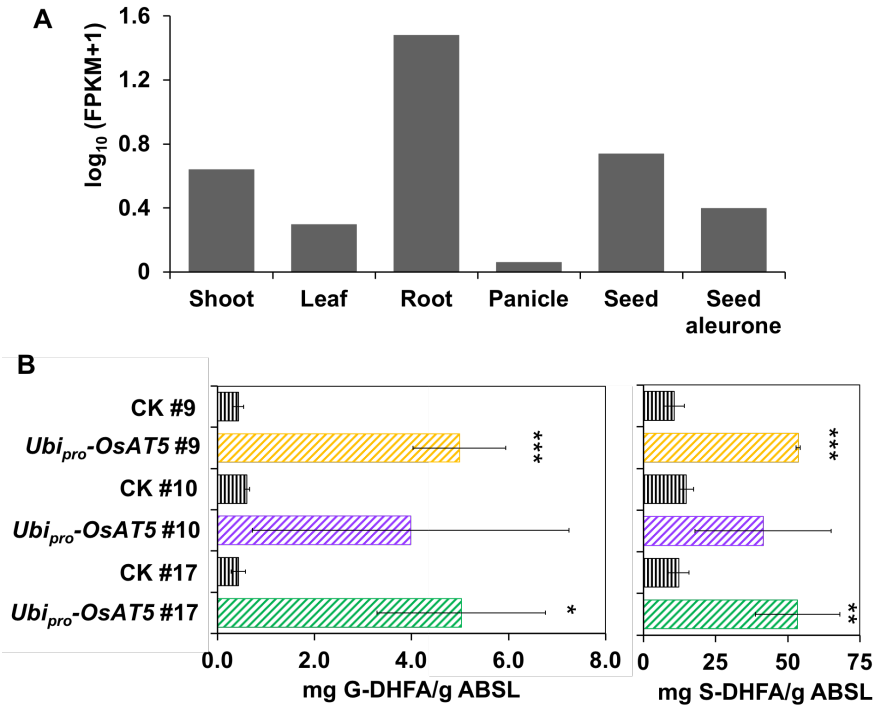


Figure 3.2 Expression pattern of *OsAT5* and effect of its overexpression on root lignin.
(**A**) *OsAT5* is most highly expressed in rice root. RNA-seq data of shoot, leaf, root, panicle, seed and seed aleurone are from International Rice Genome Sequencing Project SRP029886, SRP047482, DRP001762, and SRP028376 in Sequence Read Archive normalized by Rice Expression Database of IC4R. FPKM, fragments per kilobase of transcript per million mapped reads. (**B)** The DFRC assay shows that the Ubi_{pro}-*OsAT5* overexpression lines increased ML-DHFAs in de-starched AIR of whole mature roots. Bars indicate 2×SE of three to six biological replicates with two technical replicates measured. Student's t test, $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$ between transgenic lines and corresponding segregant wild-type (CK).

3.3.3 Increased lignin ferulate did not enhance the digestibility of rice straw

In transgenic poplar, feruloylation of lignin (called "Zip-lignin") enhances biomass deconstruction by 85% under alkaline pretreatment (Wilkerson et al., 2014). With the increased feruloylation in the rice *OsAT5* overexpression lines, we hypothesized that resulting plant biomass would be less recalcitrant and thereby more digestible. Biomass digestibility is indicated by enzyme-based glucose release from pretreated biomass. However, the yield of released glucose was not improved by *OsAT5* overexpression (*AT5*-

D1 lines), irrespective of base or acid pretreatment applied (**Figure 3.3**). Its rice straw of the transgenic may even possess a slightly decrease in glucose release, by between 4.5% and 11.6% under mild alkaline pretreatment conditions (1.5 mM 60˚C 3 hrs, 1.5 mM 90˚C 3 hrs, and 6.25 mM 90˚C 3 hrs) (**Figure 3.3**).

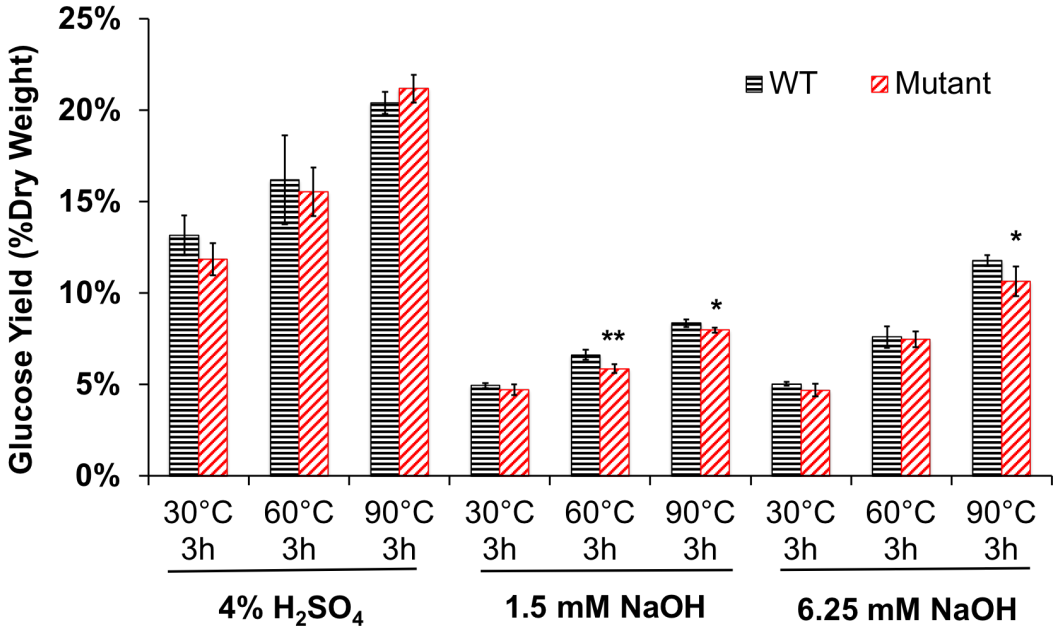


Figure 3.3 Enzymatic hydrolysis of acid- and base-pretreated biomass to glucose. Glucose yield was determined after enzymatic hydrolysis of pretreated straw from *OsAT5-D1* overexpression lines (red, *2A-20021.10.7.45)* and isogenic wild type (black, *2A-20021.11.2.40*). Data are mean ± SD of three technical replicates (ANOVA test, **P* < 0.05, ***P* < 0.01).

3.3.4 Expression of AtC4H-driven OsAT5/OsFMT in Arabidopsis successfully incorporate ML-FAs to lignin and improved the digestibility of Arabidopsis stem

To probe the impact of ML-FA conjugates on lignification and biomass digestibility in plants that do not inherently possess ML-FAs, we introduced *OsAT5* into *Arabidopsis*, which has no orthologous genes of *OsAT5* and no DFRC-detectable feruloylation in its lignin (Karlen et al., 2016). OsAT4 was included as the positive control since its

transgenic *Arabidopsis* using the secondary cell wall promoter CESA7, yielded *p*CA in lignin up to ~2 μg/mg cell wall (Smith et al., 2015). Here, we generated transgenic *Arabidopsis* for both *OsAT5* and *OsAT4* controlled either by the "constitutive" viral promoter *35S* or by the secondary cell wall promoter of the *AtC4H* gene which is active when secondary cell wall starts to grow (Yang et al., 2013). The $C4H_{pro}$-*AT4* transgenic plants had a total of 5 μg/mg *p*CA in alcohol insoluble residue (AIR) (**Supplemental Table S3.1**). To further determine which cell wall fraction is altered, we treated AIR from mature inflorescence stem with a mild, 50 mm trifluoroacetate (TFA) to cleave acid-labile glycosidic bonds (Bartley et al., 2013). In comparison, the $C4H_{pro}$-*OsAT4* lines incorporated 1.2 μg/mg of *p*CA into lignin (detected in the TFA-fraction pellet), 10 times more than wild type (0.12 μg/mg) (**Supplemental Table S3.1**).

Then, we biochemically confirmed the formation of feruloyl esters in our cell wall composition analysis with both HPLC and DFRC. Surprisingly, the $C4H_{pro}$-*OsAT5* lines boosted wall associated FA up to 7 μg/mg in the T2 generation **(Figure 3.4A),** which is even higher than normally observed in most grasses, e.g. 2.4 μg/mg FA in rice (Nipponbare) and 2.97 μg/mg switchgrass (AP13)**;** whereas FA in wildtype *Arabidopsis* inflorescence stem was not detected (the limit of *p*CA and FA detection by HPLC is at 0.01 μg/mg). Moreover, lignin-associated FA in three $C4H_{pro}$-*AT5* lines reached up to 4 μg/mg **(Figure 3.4B)** versus non-detectable in the wild type TFA-fraction. It is notable that the 35S promoter-driven *OsAT5* lines presented a much lower level of lignin feruloylation at 0.016 ± 0.005 μg/mg in AIR (**Supplementary Table S3.2**). Such a modest hydroxycinnamate level is similar to observations in transgenic *Arabidopsis* of

OsPMT, *BdPMT1*, and *BdPMT2* under the control of 35S and ZmUbi promoters (Smith et al., 2015; Sibout et al., 2016).

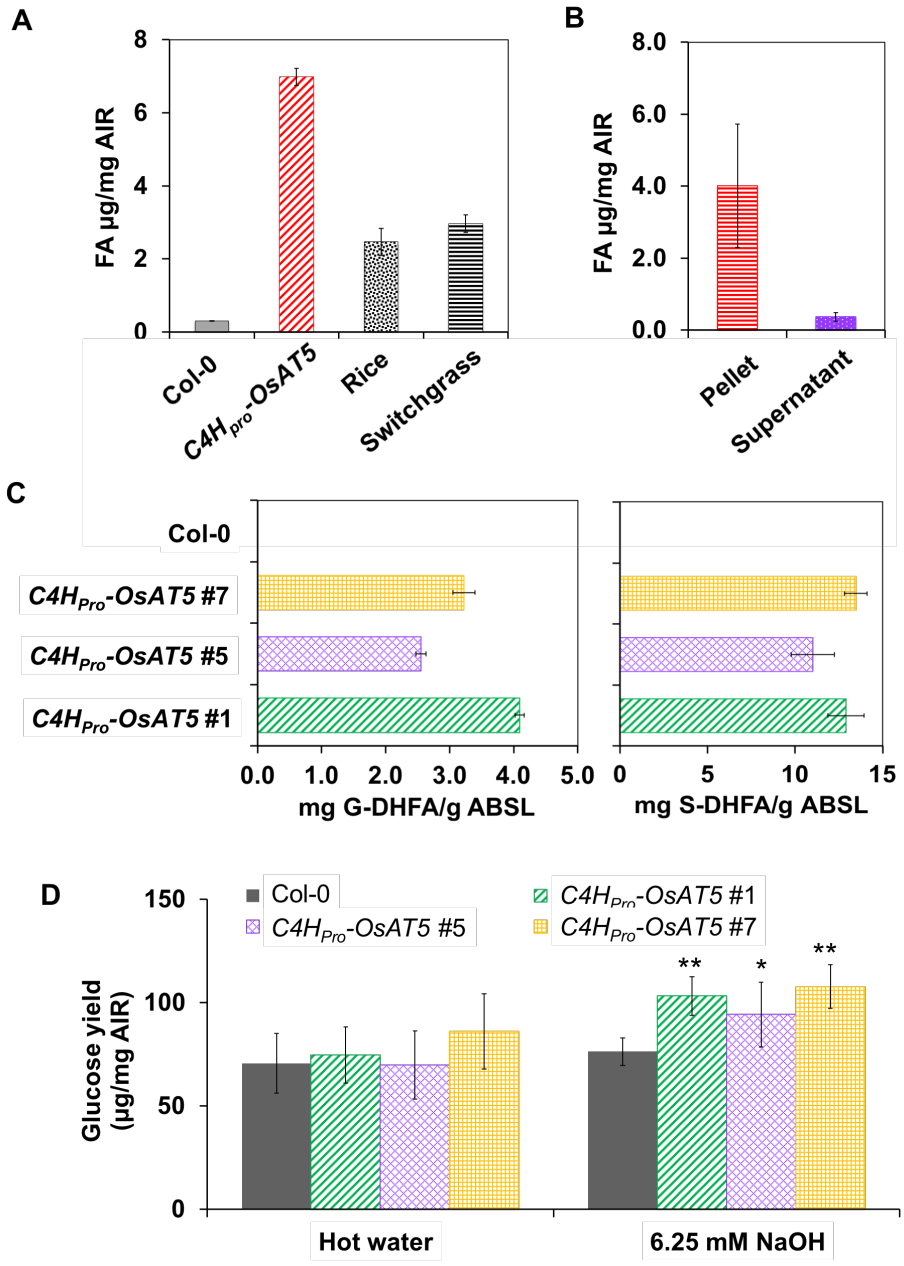


Figure 3.4 Engraftment of *OsAT5*-dependent ML-FA conjugates in *Arabidopsis* lignin improves base-pretreated biomass digestibility.
(**A)** The level of ferulic acid released by saponification from the *C4H$_{pro}$-AT5* transgenic *Arabidopsis* exceeds representative grasses. Samples are AIR of *Arabidopsis* mature inflorescence stem, mature straw of rice (Nipponbare) and switchgrass (AP13). (**B)** The increased FA in *C4H$_{pro}$-AT5* lines predominantly comes from the TFA-insoluble fraction (50 mM TFA, 100˚C 4 hrs). Both TFA-insoluble (pellet) and TFA-soluble (supernatant)

fractions were measured for T1 lines. Bars indicate 2×SE of 6 biological replicates. **(C)** The DFRC assay shows that the *C4H$_{pro}$-OsAT5* lines at T2 incorporated ML-DHFAs in influoresence stem. Bars indicate 2×SE of two technical replicates. (**D)** Glucose measurements show that the AIR fraction from mature inflorescence of three independent *C4H$_{pro}$-OsAT5* lines at T2 (#1, #5, and #7) yielded more glucose compared to the wildtype control (Col-0) after 6.25 mM NaOH 90˚C 3 hrs pretreatment, rather than hot pretreatment at 100℃ for 1 hr at pH 5. Error bars indicate 2×SE of three technical replicates (*$P < 0.05$, **$P < 0.01$, by Student's t test).

Further analysis to lignin composition by DFRC also supported that the expression of *OsAT5* successfully incorporated ML-FAs to *Arabidopsis* lignin and affected lignin polymerization. More specially, the lignin of *C4H$_{pro}$-AT5* transgenic lines gained 12.5 ± 1.5 mg/g of S-FA monolignols and 3.3 ± 0.9 mg/g of G-FA, at the S-FA/G-FA ratio of 3.9 ± 0.7 (**Figure 3.4C**); however, these conjugates were barely detected in wild type (Col-0). We also observed a small amount of ML-*p*CA in cell wall of *C4H$_{pro}$-AT5* transgenic *Arabidopsis*, with 0.6 ± 0.1 mg/g S-*p*CA lignin and 0.021± 0.006 mg/g G-*p*CA lignin (**Supplementary Table S3.3)**, which were not detected in our aforementioned yeast co-expression system and significantly changed in its overexpression in rice (Karlen et al., 2016).

To examine the recalcitrance or digestibility of OsAT5 transgenic Arabidopsis, we conducted enzyme hydrolysis for hot liquid (100 mM citrate buffer) and mild base pretreated mature inflorescence stem. As expected, the *C4H$_{pro}$-OsAT5* transgenic *Arabidopsis*, which demonstrated a higher level of ML-FAs than representative grasses (rice and switchgrass), increased the glucose release by 23-41% after 6.25 mM NaOH pretreatment compared to wild type, indicating a significant improvement in biomass digestibility. However, hot water pretreatment on the same biomass did not exhibit such

a notable change in glucose release (**Figure 3.4D**). This pretreatment tied effect on digestibility implies a zip-lignin effect posed by OsAT5 in *Arabidopsis*.

3.3.5 Differential impacts on lignin incorporation between rice Ubi_{pro}-OsAT5 *and Arabidopsis* $C4H_{pro}$-OsAT5 transgenes

OsAT5 overexpression in rice shown by the transgenic *Ubi_{pro}-OsAT5*, resulted in a 5-fold increase in both S-FA and G-FA in straw; however, an uneven increase was found in rice root, where S-FA and G-FA increased by 4-fold and 10-fold, respectively. Across plant species, the *$C4H_{pro}$-OsAT5 Arabidopsis* incorporated more S-FA (12.5 ± 1.5 mg/g) than G-FA (3.3 ± 0.9 mg/g) to its lignin. To further understand lignin feruloylation behind these observations, we examined lignin content and lignin structure in terms of the H:G:S ratio and the percentage of each type of ML-FA conjugates in all materials we examined above (**Table 3.1**). Measurements of acetyl bromide soluble lignin (ABSL) indicative of lignin content revealed no significant change in rice straw and root between wild type and *Ubi_{pro}-OsAT5* transgenes. However, the total ABSL in the *$C4H_{pro}$-OsAT5* Arabidopsis significantly decreased by 14% ($p < 0.001$), from 0.168 mg/mg dsAIR in wildtype Col-0 to 0.145 mg/mg. These results suggest that OsAT5 interfered lignin biosynthesis or incorporation in Arabidopsis rather than in rice.

Then, we measured the β-O-4 bond-released monomers, H, G, and S, and their ester conjugates, S-FA, S-*p*CA, G-FA, and G-*p*CA via DFRC (**Table 3.1**). Comparisons between non-transgenic materials, including rice straw, rice root, and Arabidopsis inflorescence stem, revealed a very little proportion of H monomer in Arabidopsis lignin and a higher proportion of H monomer in rice (5% in straw, 3% in root) but still at a low content compared to G and S. There is a larger amount of G-OH than S-OH in rice straw

(H:G:S=5:72:24) and Arabidopsis stem (H:G:S=0.8:70:30), and a similar amount of G-OH and S-OH in rice root (H:G:S=3:53:44). Considering all H, G, and S units in the DFRC-detected lignin, which includes ML-FAs, we found that wild type rice roots have more S than G (H:G:S=2:40:58), due to a larger amount of S conjugates compared to G conjugates. In wild type rice, there are very small fractions of ML-FA conjugates in both lignin polymers of straw and root. But they all have more S-FA than G-FA detected by DFRC. Among the S units, rice straw has relatively less S-FA (1.7%) than root (2.5%). Similarly, among the detected G units, G-FA accounts for 0.16% in root G lignin, two times higher than that in straw (0.08%). We did not detect any ML-FAs from Arabidopsis stem. When comparing *C4H$_{pro}$-OsAT5* transgenic lines and their corresponding wild type, we found *Arabidopsis* stem lignin dramatically decreased S units, irrespective of monomers (30% in wild type versus 14-19% in transgenic lines) and total units (30% in wildtype versus 13-19% in transgenic lines). This data indicates that expression of OsAT5 in *Arabidopsis* interfered the incorporation of S lignin. However, S-FA detected in the *C4H$_{pro}$-OsAT5* transgenic Arabidopsis accounted for 15-25% of S units, which is way more than incorporated G-FA (~1% of G units) in lignin we reported previously. It is obvious that rice root from *Ubi$_{pro}$-OsAT5* and Arabidopsis stem from *C4H$_{pro}$-OsAT5* presented different impacts on lignin composition, as a result of OsAT5 overexpression.

Table 3.1 Effect of *OsAT5* overexpression on lignin structure in rice straw, rice root and *Arabidopsis* inflorescence stem.

Genotype	Tissue	H:G:S	Total H:G:S [1]	ABSL (μg/mg)	S-FA% [2]	G-FA% [3]
Rice WT	Straw	5:72:24	4:61:35	0.10±0.01[a]	1.7±0.7%[a]	0.08±0.01%[a]
Ubi$_{pro}$-OsAT5	Straw	6:78:17	5:62:34	0.10±0.01[a]	11±3%[b]	0.5±0.1%[a]
Rice WT	Root	3:53:44	2:40:58	0.22±0.01[b]	2.5±0.1%[a]	0.16±0.04%[a]
Ubi$_{pro}$-OsAT5	Root	6:56:38	3:36:61	0.21±0.01[b]	13.7±3%[bc]	2.5±0.4%[b]
Col-0	Stem	0.8:70:30	0.75:70:30	0.168±0.002[c]	0[a]	0[a]
C4H-OsAT5	Stem	0.7:82:17	0.7:77:16	0.145±0.005[c]	21±6%[c]	1.5±0.4%[c]

[1]Total H = HOH, total G = GOH + G-DHFA + G-DH *p*CA, total S = SOH + S-DHFA + S-DH *p*CA;
[2]S-FA% = S-FA/ (SOH + S-DHFA + S-DH *p*CA), the percentage of ML-FA conjugates in total ML;
[3] G-FA % = G-FA/ (GOH + G-DHFA + G-DH *p*CA).
Data are mean ± 2x SE of three rice transgenic lines with 3-6 biological replicates and two technical replicates, *Arabidopsis* lines with 40 plants pooled for measurements with two technical replicates (ANOVA with Tukey test within each column, $p<0.05$ was denoted by different groups i.e. a, b, c).

3.3.6 Lignin is more soluble in rice than in Arabidopsis *in alkaline medium*

Since ML-FAs are easily cleavable during alkaline treatment, the question we raised here is whether the increased level of ML-FAs in our overexpression mutants will improve lignin solubility in alkaline. We treated dsAIR with 1 M NaOH for 15 hrs and precisely weighted the starting materials and treated materials, then compared lignin loss in the soluble supernatants between samples by simply subtracting ABSL of treated residues from the starting materials. We found that the *C4H$_{pro}$-OsAT5* Arabidopsis had 60% lignin solubilized in alkaline, which is a significant increase compared to 20% in wildtype (Col-0). In contrast, the *Ubi$_{pro}$-OsAT5* transgenic rice did not change alkaline-soluble lignin in root. With comparisons among different wildtype plant species under the same test

condition, we found that 73% ± 4% of lignin in both rice straw and root was solubilized in alkaline, 59% ± 5% in switchgrass straw, 50% ± 3% in Arabidopsis stem, and 47%± 2% in Poplar stem (**Figure 3.5**). These results suggest that rice lignin is more likely to be removed by alkaline and lignin might be a less important recalcitrance factor in rice than in other plant species.

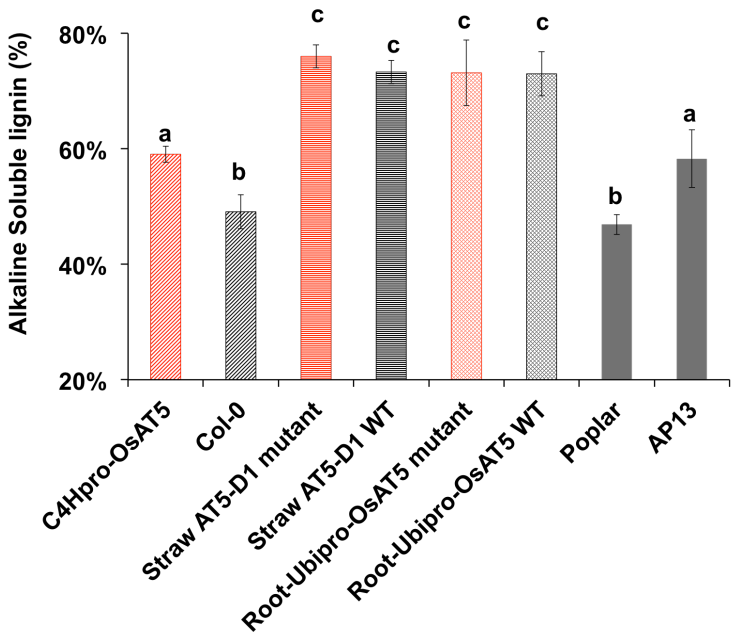


Figure 3.5 Measurement of alkaline soluble lignin in plant species.
Materials include stem of poplar, straw of switchgrass (AP13), straw of rice *AT5-D1* mutant and wildtype (Dongjin), root of rice *Ubi$_{pro}$-OsAT5* mutant and wildtype (Nipponbare), *Arabidopsis C4H$_{pro}$-OsAT5* mutant and wildtype (Col-0). Values represent percentages of ABSL solubilized in 1 M NaOH at room temperature (15 hrs) to the total ABSL of dsAIR cell wall materials. The means and 2xSE were calculated from the values of three technical replicates of switchgrass, poplar, Col-0 and straw of *AT5-D1*, and three biological replicates of root materials of *Ubi$_{pro}$-OsAT5* rice, and *C4H$_{pro}$-OsAT5 Arabidopsis*. ANOVA with Tukey test, $p<0.05$ was shown by different group, i.e. a, b, c.

3.4 Discussion

Monolignol ferulates are naturally incorporated into plant lignin, representing a newly recognized monomer., This modification has been hypothesized to alter alkaline-base

treatment of plant biomass due to the chemical features of ester bonds. Here we characterized a rice feruloyl monolignol transferase (OsAT5/OsFMT), which can synthesize ML-FA conjugates. However, plants have convergently evolved this biofunction since most grasses have both S-FA and G-FA but dicots like poplar only have G-FA in lignin. Most likely, they are using different enzymes in clades of the BAHD acyltransferase family (Karlen et al., 2016). It is still uncertain how this native OsFMT alters lignin structure and its relationship with biomass recalcitrance. Our results verified the enzymatic function of OsAT5, but also showed that heterologous expression of OsAT5 in Arabidopsis, but not in rice, improves biomass digestibility. Our results also provided possible explanations for this observed difference.

Heterologous expression of OsAT5 in yeast with feeding putative substrates provides strong evidence that OsAT5 possesses FMT activity; however, substrate preference could not be conclusively determined. Yeast has been previously demonstrated to be an amendable host for producing monolignol hydroxycinnamates by co-expressing the BAHD acyltransferase and Arabidopsis At4CL5 (Eudes et al., 2016c). This system also confirmed the PMT activity of the well-studied OsAT4, which has the highest affinity in vitro for *p*-coumaryl alcohol (H) and *p*CA -CoA (Withers et al., 2012). In our yeast experiments upon feeding with *p*CA-CoA and coniferyl alcohol, OsAT4 produced coniferyl coumarate (G-*p*CA), which also accumulated with expression of OsAT4 in Arabidopsis and poplar (Smith et al., 2015). However, we did not observe the sinapyl coumarate transferase activity of OsAT4 in the yeast system, though this had been previously observed in enzyme kinetic analysis and genetics (Withers et al., 2012; Smith et al., 2015). In addition, OsAT4 in yeast showed weak production of coniferyl ferulate

with feeding of FA-CoA and coniferyl alcohol, the activity of which has not been reported for OsAT4 but has been shown by its close homolog BdPMT1 (OsAT3) from Brachypodium when expressed in the *ccrg1* Arabidopsis mutant. Altogether, the yeast system seems inefficient at synthesis of sinaypl hydroxycinnamates (S-conjugates) while favoring coniferyl hydroxycinnamates (G-conjugates). From this point of view, it is not surprising that the yeast expression failed to demonstrate the sinapyl feruloyl transferase activity observed in overexpression of OsAT5 in rice (*Ubipro-OsAT5*) and Arabidopsis (*C4Hpro-OsAT5*).

The cell wall accumulation of monolignol-FA conjugates in *$C4H_{pro}$-OsAT5* Arabidopsis adds to other data showing that *Arabidopsis* and other dicots can transport these conjugates to the cell wall for polymerization. The *$C4H_{pro}$-OsAT5* lines successfully incorporated ML-FA and ML- *p*CA conjugates to Arabidopsis lignin, which is consistent with previous reports in *CESA7::AsFMT* poplar lines (Wilkerson et al., 2014), proAtPRX64::FMT (Smith et al., 2017b), *CESA7::OsAT4* Arabidopsis lines (Smith et al., 2015), and *C4H-BdPMT1/BdPMT2* Arabidopsis lines (Sibout et al., 2016). Poplar lignin contains G-FA conjugates, but no conjugates could be detected in wild-type Arabidopsis cell walls within the limit of detection (Karlen et al., 2016). Thus, the plasticity of lignin polymerization shown in these studies suggests that plants may share similar mechanisms for monolignol transportation and polymerization.

Using a temporal and tissue-specific promoter appears to be important for the study of cell wall biosynthesis genes. Consistent with many reports, our study also demonstrated that secondary cell wall promoter-driven expression resulted in a more obvious cell wall phenotype compared to constitutive promoters e.g. 35S and ZmUbi

(Smith et al., 2015; Sibout et al., 2016; Smith et al., 2017b; Aznar et al., 2018). In our case, the dramatic increase in HCA shown in the *C4H* promoter-driven mutants, compared to the *35S$_{pro}$-AT5* plants, may be due to the difference in cell wall composition and structure in different cell types and tissues. This promoter issue may also contribute to the subtle effect on enzyme digestibility of the *Ubi$_{pro}$-OsAT5* lines.

The effect of OsAT5 on lignin structure varies with plant species and organs. Both rice and Arabidopsis transgenic lines were able to synthesize S-FA and G-FA and incorporate them into lignin. However, overexpression of OsAT5 in rice root showed more G-FA incorporated into lignin, whereas *C4H$_{pro}$-OsAT5* Arabidopsis stem conversely prefers to incorporate S-FA. After deep examination on lignin content and structure in this study, we found that the lignin of wild type rice root has more acylated S units (S-FA and S- *p*CA) than acylated G units (G-FA and G- *p*CA), resulting in an opposite S:G ratio pattern from rice straw and Arabidopsis Col-0. Overexpression of OsAT5 rendered the rice root incorporated more G-FA into lignin but did not significantly change the S:G ratio (**Table 3.1**). The S:G ratio was only significantly decreased in the *C4H$_{pro}$-OsAT5* lines compared to Col-0, where more S-FA was incorporated to Arabidopsis stem lignin but less S monomers incorporated via β-O-4 bonds (DFRC based). In addition, the lignin content was significantly decreased in the *C4H$_{pro}$-OsAT5* lines. This may indicate the decreasing trend on the biosynthesis or incorporation of S monomers. However, we cannot exclude the limitation of the DFRC method that only cleaves β-O-4 bonds in the lignin polymer (Lu and Ralph, 1997). As the incorporated ML-FA conjugates also can undergo cross coupling by stronger ether bonds β-O-4′, β-5′, α-O-4′, 4-O-5′, β-β′ in common, and β-1′, and 5-5′ in minor amount (i.e. 4-O-5', α-O-4')

or C-C bonds (i.e. β-β', β-5', β-1', 5-5'). This effect is unseen based on the current lignin analyzing approach. However, we observed more diferulates from *AT5-D1* straw organs (**Supplementary Figure S3.4**). In consideration of this limitation, we cannot conclude whether the differential lignin context could exert any effects on substrate availability for OsAT5.

The contrasting deconstruction results upon expression of OsAT5 in Arabidopsis and rice suggest that cell wall recalcitrance is affected by the incorporation of ML-Fas. Differences may depend on lignin content and structure, as well as other unseen changes in cell wall polymer associations. Lignin content and structure are two important factors influencing cell wall recalcitrance. Overexpression of OsAT5 in rice did not improve, even slightly decreased the enzymatic digestibility after mild base pretreatment. The lignin content and structure, in terms of H:S:G ratio, did not change in these lines. Besides, more ML-FA conjugates incorporated in Ubi_{pro}-*OsAT5* lines (compared to wild type) did not affect the alkaline lignin solubility. But the digestibility of cell wall of $C4H_{pro}$-*OsAT5* Arabidopsis lines was greatly improved after mild base pretreatment. This can be caused by the decrease in lignin content and the S:G ratio. These changes on lignin also improved lignin solubility in alkaline. But, why did the increased ML-FA conjugates have different effects on lignin alkaline solubility? When we compared the lignin solubility of representative wild-type dicots (Arabidopsis and poplar) and grasses (rice and switchgrass), we found that rice lignin is the most soluble in mild alkaline, switchgrass as the second, Arabidopsis and poplar are the least (**Figure 3.5**), suggesting that lignin in rice are smaller than other species examined here. Thus, we deduce that the increasing ML-FA conjugates may actually interconnect lignin polymers together in the Ubi_{pro}-

OsAT5 lines, but since the change is relatively low probably due to the insufficient overexpression effect (promoter issue), and hidden in the intrinsic alkaline solubility feature of rice lignin. On the other hand, ML-FA conjugates in rice lignin are also capable of cross coupling the ferulate on arabinoxylan, which is featured in grasses, thereby enhancing the crosslinking between lignin and arabinoxylan. This also can explain why more ML-FA conjugates in rice lignin still decreased enzymatic digestibility a bit. Nonetheless, it also suggests that manipulating lignin content and structure is less important than what we expected in rice and other grasses. More and more studies support that lignin is less important than hemicellulose in contributing to grass recalcitrance. DeMartini et al. compared the digestibility of switchgrass (grass) and poplar (dicot wood) under the same procedure of removing specific polymers from cell wall and found that hemicellulose remarkably contributes to grass recalcitrance (DeMartini et al., 2013). A genome-wide association study for lignin abundance and sugar yield in maize also supported this notion (Penning et al., 2014). In addition, tissues also have different contributors for recalcitrance. For example, enzyme hydrolysis of switchgrass stem is most affected by lignin; however, the hemicellulose abundance influences more in hydrolysis of leaf materials (Crowe et al., 2017). Our work not only provides a deeper understanding of biomass recalcitrance, in which ML-FA conjugates can play both negative and positive roles, but also form a new strategy for the engineering of eudicot energy crops.

3.5 Materials and Methods

3.5.1 Transgenic plants and growth condition

We amplified the open reading frame of OsAt4 (LOC_Os.1) from Nipponbare seedling complementary DNA (cDNA) with primers listed in Supplemental Table S3. The full coding sequence of OsAT5 has been synthesized (Karlen et al., 2016). Both genes were integrated into the pAtC4H-GW-Nos backbone to generate *$C4H_{pro}$-OsAT5* and *$C4H_{pro}$-OsAT4* constructs, in which genes are sandwiched by the AtC4H promoter and the 3'-terminator of nopaline synthase. We also cloned these genes into the p35S-GFP-GW-Nos backbone to produce *$35S_{pro}$-OsAT5* and *$35S_{pro}$-OsAT4* constructs.

Transgenic Arabidopsis were generated by floral dip into *Agrobacterium tumefaciens* (Clough and Bent, 1998). Seeds were sterilized and plated on ½ Murashige and Skoog (MS) plates with 20 mg /L hygromycin for *$C4H_{pro}$-OsAT5* and *$C4H_{pro}$-OsAT4* or with kanamycin 100 mg/L for *35S::OsAT5* and *35S::OsAT4*, and then kept in the dark at 4°C for 2 days for cold stratification. Seven days old seedlings were transplanted from plates to soil. The type of soil and growth condition have been reported in Jones et al. (Jones et al., 2017). All transgenic plants were confirmed by genotyping and qRT-PCR. Primers are listed in **supplemental Table 3.3**.

3.5.2 Yeast feeding experiment

Vector Construction

OsAT4 and *OsAT5* were cloned into the *Saccharomycetes cerevisiae* expression vector pDRf1-4CL5-GW (Eudes et al., 2015) using LR clonase II, allowing constitutive co-expression with *Arabidopsis thaliana* 4-coumaryl:CoA ligase (At4CL5) in yeast. The resultant pDRf1-4CL5-OsAT4 and pDRf1-4CL5-OsAT5 constructs were transformed

into *S. cerevisiae* pad1 knockout (*MAT*a *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0* Δ*pad1*, ATCC 4005833) (Winzeler et al., 1999) using the Frozen-EZ Yeast Transformation II Kit and selected on uracil dropout solid medium.

Acyltransferase Activity Assay

Yeast colonies harboring pDRf1-4CL5-OsAT4 or pDR1-4CL5-OsAT5 were inoculated in 2X yeast nitrogen base (YNB) without amino acids but supplemented with 3% glucose and 2X dropout-uracil (CSM-ura), then grown at 30°C, 200 rpm, 16 hrs. The overnight culture after diluted to OD_{600} of 0.1 to 0.2 continued to grow at 30°C, 200 rpm until OD_{600} reached 1. Cinnamic acids and cinnamoyl alcohols after suspended in DMSO were added to the yeast culture at 400 μM. The fed cultures were further cultured for 24 hrs with shaking at 200 rpm, 30°C. Cell pellet was spinned down at 20,800 x *g* for 3 min and then stored for analysis. Control yeast harboring the pDRf1-4CL5-GW vector was also sampled in parallel.

Metabolite Extraction

Metabolite extraction was previously described (Eudes et al., 2016a). Briefly, 1 mL cyclohexane : water (1:1, v/v) was added to each Yeast cell pellet and vortexed for 1 min. The mixture was centrifuged (20,800 x g, 5 min, 4°C) and the upper organic phase was collected and evaporated under nitrogen using a sample concentrator. The dried metabolite was resuspended in 100 μL water : methanol (1:1, v/v) prior to LC-MS analysis.

LC-MS analysis

Metabolites were separated by an Agilent Technologies 1100 Series HPLC system using the Kinetex XB-C18 column (100-mm length, 3.0-mm diameter, and 2.6

μm particle size). The sample tray and column were set to 4°C and 50°C, respectively. 5 μL of samples was injected and then eluted isocratically with 90% methanol at the flow rate of 0.42 mL/min. The HPLC system was coupled to an Agilent Technologies 6210 series time-of-flight mass spectrometer (for LC-TOF MS) via a MassHunter workstation.

3.5.3 Cell Wall Analyses

Preparation of AIR (Alcohol insoluble residue) and dsAIR (de-starched AIR)

Both root materials of *Ubi$_{pro}$-OsAT5* transgenic plants and wildtype were harvested at the post-reproduction stage. At the same time, we also harvested straw (Karlen et al., 2016). Root tissues after washed with water were dried at 45˚C for 7 days. The mature inflorescence stems of Arabidopsis transgenic plants and control (Col-0) were collected in two weeks after stopping watering and at that time, plant materials were totally dry. All biomass materials were ground by shaking with a tissue homogenizer in 2-mL polypropylene tubes at 1,400 rpm with one big stainless-steel ball and four small balls for 2 min. To make AIR, ground materials were subjected to multiple ethanol hot washes as described previously (Bartley et al., 2013) and then lyophilized afterward. The de-starched materials were obtained by treating AIR with amylase and amyloglucosidase (Bartley et al., 2013).

Analysis of Hydroxycinnamic Acids

HCA analysis was conducted as described previously but with minor changes (Bartley et al., 2013). AIR or dsAIR samples were saponified with 2N NaOH for 24 hrs at room temperature and transcinnamic acid was added as an internal control. Then, the extracts were neutralized with concentrated HCl, diluted 1:1 with nanopure water, and

then subjected to high-performance liquid chromatography (HPLC) analysis as reported before.

Hydroxycinnamate Fractionation

To run trifluoroacetic acid (TFA) fractionation assays, 2.5 mg of AIR or dsAIR samples were mixed with 500 µl of 0.05 M TFA, followed wtih heating at 100˚C for 4 hr. Then, supernatant and pellet were treated separately with 2M NaOH for HCA extraction. Transcinnamic acid was added as internal control for both supernatant and pellet fractions. Extracts were then subject to HPLC analysis as previously described.

Dimers

Drs. J. Ralph, S. Karlen, F. Lu, and R. Gao provided standards of ferulate dehydrodimers (Ralph et al., 1994). We used meta-hydroxycinnamic acid (mCA) for preparation of the calibration standards. 5 mg of dsAIR was used for HCA extraction and then analyzed by HPLC.

DFRC

The DFRC analysis was performed as described by Smith et al. (2017). Briefly, Arabidopsis AIR samples (50 mg) were treated with a solution of acetyl bromide (AcBr) in acetic acid (20%, v/v) at 50 °C for 2.5 h in a 3-dram vial with a vacuum-sealed cap. The acetylated and benzyl-brominated lignin solution was dried under vacuum for 35 min at 50 °C. The dry film was treated with absolute ethanol (1 mL), which was then also removed on a SpeedVac concentrator at 50 °C for 15 min. The dry sample was then immediately dissolved in a mixture of 1,4-dioxane:acetic acid:water (5:4:1, by volume, 5 mL), and zinc nanopowder (150 mg) was added to the vial. The reaction was stirred for 2 h at room temperature and then quenched with saturated ammonium chloride. The

quenched reaction was spiked with an internal standard mixture (deuterated monolignols and conjugates). The organics were extracted with DCM (3 × 15 mL) and the combined organic fractions were dried over sodium sulfate. The DCM was removed under vacuum, and the free hydroxyl groups were acetylated overnight using a mixture of acetic anhydride and pyridine (1:1, v/v). The excess acetic anhydride and pyridine were removed on a rotary evaporator, after which the crude product was loaded onto a Supelco Supelclean LC-SI SPE tube (Sigma-Aldrich) with the aid of ethyl acetate and hexane (1:1, ~1 mL). The purified product was then eluted using a mixture of ethyl acetate:hexane (1:1, 10 mL), and concentrated to dryness. The dry film was dissolved in DCM (1 mL) and injected into a Shimadzu GCMS-TQ8030 triple-quadrupole GC/MS/MS operating in multiple-reaction-monitoring (MRM) mode for quantitative analysis (using calibration curves derived from synthetic standards) of monolignols. The samples were then concentrated under vacuum (SpeedVac) to 200 uL and injected into the GC/MS/MS again for quantitative analysis of monolignol ferulates and monolignol *p*-coumarates. Each line represents an average of two technical replicates.

Acetyl bromide Lignin

Lignin was quantified by acetyl bromide solubilization (Fukushima and Hatfield, 2004), then quantified by measuring the absorbance at 280 nm in a UV-compatible 96-well plate. We followed the same procedure as reported previously (Bartley et al., 2013).

3.5.4 Alkaline solubility lignin assay

About 50 (±5) mg of dsAIR samples was put into a 15-mL falcon tube together with 10 mL of 1 M NaOH. The empty tube and samples were both measured. The suspension was agitated at room temperature for 15 hrs, then neutralized with 1 ml of

10M HCl. After carefully removal of the supernatant, the residues were washed with di-water 3 times and then lyophilized. The mass of sponified residues was measured by subtracting the empty tube. Then the ABSL were measured for both starting materials and sponified residues and converted to total mass of each. The soluble alkaline lignin percentage was calculated by the following formula, $ABSL\%=[ABSL_{dsAIR}*W_{dsAIR}-ABSL_{residue}*W_{residue}]\ \%/(\ ABSL_{dsAIR}*W_{dsAIR})$

3.5.5 Enzymatic Saccharification Assay

AIR (2-5 mg) of Arabidopsis stem samples was pretreated by shaking at 90˚C in 500 µl of 100 mM citrate buffer (pH 5.0) for 3 hrs and 6.25 mM NaOH 90˚C 1 hr, respectively. Base pretreated samples were then neutralized. Pretreated samples were then digested with a mix of NS50013 which contains a cellulase cocktail, and NS50010, which contains beta-glucosidase, t at 50°C with mild shaking. Released reducing sugars were quantified by the 3,5-dinitrosalicilate (DNS) assay (Bartley et al., 2013).

Rice enzymatic digestion assay was done by iWALL (Automated Grinding, Feeding, and Weighing System) from Department of Engergy Great Lakes Bioenergy Research center at Michigan State University. In brief, Accellerase 1000 in 30 mM citrate buffer (pH 4.5) plus 0.01% sodium azide was used to hydrolyze materials from different pretreatments. Then, glucose was assayed with the glucose oxidase/peroxidase (GOPOD) method.

Acknowledgments

We thank Dr. Dominique Loque for the secondary cell wall binary construct. MLP, CZ, and LEB were funded in part by the DOE Feedstock Genomics (DOE BER Office of Science DE-SC006904), in part by the National Science Foundation EPSCoR program

under Grant No. EPS-0814361, and in part by the University of Oklahoma College of Arts and Sciences. MFL was supported by the University of Oklahoma honors research program SDK, RAS, and JR were funded by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494).

3.6 References

Aznar A, Chalvin C, Shih PM, Maimann M, Ebert B, Birdseye DS, Loqué D, Scheller H V. (2018) Gene stacking of multiple traits for high yield of fermentable sugars in plant biomass. Biotechnol Biofuels 11: 2

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, et al (2013) Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. PLANT Physiol 161: 1615–1633

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54: 519–546

Bunzel M, Ralph J, Lu F, Hatfield RD, Steinhart H (2004) Lignins and Ferulate−Coniferyl Alcohol Cross-Coupling Products in Cereal Grains. J Agric Food Chem 52: 6496–6502

Clough SJ, Bent AF (1998) Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. Plant J 16: 735–743

Crowe JD, Feringa N, Pattathil S, Merritt B, Foster C, Dines D, Ong RG, Hodge DB (2017) Identification of developmental stage and anatomical fraction contributions to cell wall recalcitrance in switchgrass. Biotechnol Biofuels 10: 673–686

DeMartini JD, Pattathil S, Miller JS, Li H, Hahn MG, Wyman CE (2013) Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. Energy Environ Sci 6: 898

Eudes A, Benites VT, Wang G, Baidoo EEK, Lee TS, Keasling JD, Loqué D (2015) Precursor-directed combinatorial biosynthesis of cinnamoyl, dihydrocinnamoyl, and benzoyl anthranilates in saccharomyces cerevisiae. PLoS One 10: e0138972

Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai Y-L, Baidoo EEK, Chiu T-Y, Heazlewood JL, et al (2016a) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. Microb Cell Fact 15: 198

Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai YL, Baidoo EEK, Chiu TY, Heazlewood JL, et al (2016b) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. Microb Cell Fact. doi: 10.1186/s12934-016-0593-5

Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai YL, Baidoo EEK, Chiu TY, Heazlewood JL, et al (2016c) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. Microb Cell Fact. doi: 10.1186/s12934-016-0593-5

Fukushima RS, Hatfield RD (2004) Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. J Agric Food Chem 52: 3713–3720

Grabber JH, Hatfield RD, Lu F, Ralph J (2008) Coniferyl ferulate incorporation into lignin enhances the alkaline delignification and enzymatic degradation of cell walls. Biomacromolecules 9: 2510–2516

Grabber JH, Hatfield RD, Ralph J (1998) Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. J Sci Food Agric 77: 193–200

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science (80-) 315: 804–807

Jones DS, Yuan J, Smith BE, Willoughby AC, Kumimoto EL (2017) MILDEW RESISTANCE LOCUS O Function in Pollen Tube Reception Is Linked to Its Oligomerization and Subcellular Distribution. 175: 172–185

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan DDD, Helmich KE, Free HCA, Lee S, Smith BG, Lu F, et al (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci. Adv. 2:

Lu F, Ralph J (1997) Derivatization Followed by Reductive Cleavage (DFRC Method), a New Method for Lignin Analysis: Protocol for Analysis of DFRC Monomers. J Agric Food Chem 45: 2590–2592

Marita JM, Hatfield RD, Rancour DM, Frost KE (2014) Identification and suppression of the *p*- coumaroyl CoA:hydroxycinnamyl alcohol transferase in *Zea mays* L. Plant J 78: 850–864

Penning BW, Sykes RW, Babcock NC, Dugard CK, Held MA, Klimek JF, Shreve JT, Fowler M, Ziebell A, Davis MF, et al (2014) Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population. Plant Physiol 165: 1475–1487

Petrik DL, Karlen SD, Cass CL, Padmakshan D, Lu F, Liu S, Le Bris P, Antelme S, Santoro N, Wilkerson CG, et al (2014) p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. Plant J 77: 713–26

Ragauskas AJ, Li M, Pu Y (2016) Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance. doi: 10.3389/fchem.2016.00045

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, et al (2006) The path forward for biofuels and biomaterials. Science (80-) 311: 484–489

Ralph J (2010) Hydroxycinnamates in lignification. Phytochem Rev 9: 65–83

Ralph J, Quideau S, Grabber JH, Hatfield RD (1994) Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. J Chem Soc Perkin Trans 1 0: 3485

Sibout R, Le Bris P, Legée F, Cézard L, Renault H, Lapierre C, Legee F, Cezard L, Renault H, Lapierre C, et al (2016) Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA:monolignol transferase (PMT). Plant Physiol 170: pp.01877.2015

Smith R a, Gonzales-Vigil E, Karlen SD, Park J-Y, Lu F, Wilkerson C, Samuels a. L, Ralph J, Mansfield SD (2015) Engineering monolignol p-coumarate conjugates into Poplar and Arabidopsis lignins. Plant Physiol 169: pp.00815.2015

Smith RA, Cass CL, Mazaheri M, Sekhon RS, Heckwolf M, Kaeppler H, de Leon N, Mansfield SD, Kaeppler SM, Sedbrook JC, et al (2017a) Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol ferulates incorporated into maize lignins. Biotechnol Biofuels 10: 109

Smith RA, Schuetz M, Karlen SD, Bird DA, Tokunaga N, Sato Y, Mansfield SD, Ralph J, Samuels AL (2017b) Defining the Diverse Cell Populations Contributing to Lignification in Arabidopsis thaliana Stems. Plant Physiol pp.00434.2017

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol 153: 895–905

Wilkerson CG, Mansfield SD, Lu F, Withers S, Park J-Y, Karlen SD, Gonzales-Vigil E, Padmakshan D, Unda F, Rencoret J, et al (2014) Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. Science 344: 90–3

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science (80-) 285: 901–6

Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG (2012) Identification of Grass-specific Enzyme That Acylates Monolignols with p-Coumarate. J Biol Chem 287: 8347–8355

Yang F, Mitra P, Zhang L, Prak L, Verhertbruggen Y, Kim J-S, Sun L, Zheng K, Tang K, Auer M, et al (2013) Engineering secondary cell wall deposition in plants. Plant Biotechnol J 11: 325–335

Chapter 4 Overexpression of BAHD acyltransferase, *OsAT9,* alters hydroxycinnamate content of rice cell walls and increases biomass recalcitrance

Author: Chengcheng Zhang, Jillian C. Vaught, Laura E. Bartley

Publication status: This chapter is planned for publication but will require further experimentation.

Author contribution: CZ and LEB designed this study. CZ cloned genes and made constructs. CZ genotyped and characterized transgenic plants, including gene expression, HCA analysis, deconstruction assays and sample prep for immunolocalization assays. CZ wrote the manuscript.

4.1 Abstract

Feruloyl arabinoxylan is a distinguishing feature of grasses and related monocots. Ferulate participates in radical coupling to form cross-links between cell wall polymers, which hamper biomass degradation during biochemical conversion to biofuels. To understand the mechanism of feruloylation in grasses, we used reverse genetics to examine the function of a so-called BAHD acyltransferase, OsAT9, in rice (*Oryza sativa*). Our results showed that overexpression of *OsAT9* with the maize Ubiquitin promoter in rice increased the ratio of ferulic acid to *p*-coumaric acid in cell wall polysaccharides, improved the extractability of xylan by 50-56% after base treatment, and reduced the enzymatic digestibility of leaves and stems. These results suggest that OsAT9 is a strong candidate for being a feruloyl arabinosyl transferase. The new understanding of feruloylation of rice cell walls revealed by this study can guide plant biomass engineering.

4.2 Introduction

Feruloyl arabinoxylan is a distinguishing cell wall feature in commelinid monocotyledons, including cereals and other grasses (Harris and Trethewey, 2010; Karlen et al., 2016). Wall-bond ferulic acid (FA) is abundant in Arecales (palms), Commenlinales (water hyacinth), Poales (grasses) and Zingiberales (banana, ginger) (Harris and Hartley, 1980; Vogel, 2008) representing up to 4% of cell wall mass in grasses (Hatfield et al., 1999; Vogel, 2008). This group of plants have xylan as the main hemicellulose component in cell walls, comprising up to 40% of dry weight biomass (Scheller and Ulvskov, 2010). Feruloylation mainly occurs on (glucurono)arabinoxylans (GAXs), and recent reports found it is also able to esterify lignin (Karlen et al., 2016). GAX-bound ferulates widely exist in commercially important species, such as barley grain and straw, wheat bran, rye, Lolium, rice, Brachypodium, maize bran, Phalaris, bamboo, oat, and pineapple (de O Buanafina, 2009).

Biochemically, ferulate participates in radical coupling to form cross-links, which correlates with limited biomass degradation in ruminant animals and also biochemical conversion of biomass for biofuel production. In grasses, most ferulate attaches to the C5-hydroxyl of arabinosyl side chains of xylan (AX) in both primary and secondary cell walls. Ferulate also forms dimers, trimers, and tetramers, consequently cross-linking major non-cellulosic polymers together. In secondary cell walls, ferulate on polysaccharides can also initiate lignin synthesis through ester-ether linkages by radical coupling, by which lignin is anchored to polysaccharides. Several in vitro enzyme digestion assays suggest that FA amounts and cell wall digestibility are negatively correlated (Iiyama et al., 1994; Grabber et al., 1998; Lam et al., 2003; Casler and Jung,

2006; Matias de Oliveira et al., 2014). FA dehydrodimer cross-linkages can impede enzyme access to cellulose (Akin et al., 1993; Wojtaszek, 1997; Damásio et al., 2013). In line with that, heterologous expression of the feruloyl esterase which enzymatically cleaves ester bonds in transgenic *Lolium* increases biomass digestibility (de O Buanafina, 2009). Thus, understanding the mechanism of feruloylation on arabinoxylan in grasses will provide valuable clues for breeding or plant engineering for a digestible biomass.

Candidate genes involved in AX feruloylation were first identified by differential gene expression between grasses and dicots (Mitchell et al., 2007). They fall into a clade V of the BAHD transferase superfamily (Tuominen et al., 2011), several of which function in decorating cell walls with hydroxycinnamates. In rice, co-suppression of OsAT7, OsAT8, OsAT9 and OsAT10 by RNA interference (RNAi) reduced cell wall FA by 20% (Piston et al., 2010). Later, OsAT10 was genetically implicated in being responsible for the acylation of AX by *p*-coumaric acid (*pCA*) (Bartley et al., 2013). Increased expression of *OsAT10* increased AX-*p*CA abundance, but also caused a partial decrease in AX feruloylation (Bartley et al., 2013). Like FA, *p*CA is a hydroxycinnamate, but *p*CA lacks the O-methyl on 3-C phenyl ring carbon. *p*CA not only extensively presents in lignin polymers from various commelinid species (Karlen et al., 2018), but also occurs on GAXs in modest amounts (Mueller-Harvey et al., 1986; Chiniquy et al., 2012; Bartley et al., 2013; Petrik et al., 2014). Other ATs that have been characterized so far conjugate hydroxycinnamates onto monolignols, which can then be incorporated onto lignin polymers. OsAT4 and OsAT3 and the Brachypodium ortholog of OsAT8 are *p*-coumaryl monolignol transferases (PMTs); AT5 is a feruloyl monolignol transferase (FMT) (Petrik et al., 2014; Smith et al., 2015b; Karlen et al., 2016; Sibout et al., 2016).

OsAT1 and OsAT9 are strong candidates for acylation of AX with FA in rice. Silencing the ortholog of OsAT1 in Brachypodium (BdAT1) showed ~25% decrease in FA-esters, in concert with increased FA-esters by *BdAT1* overexpression (Buanafina et al., 2015). The mild changes in FA-esters in *Ubi:RNAi:BdAT1* and *35S:BdAT1*, on the one hand, may be caused by the promoters used. On the other hand, there may be compensatory mechanisms or gene redundancy within this clade. Within the 10 members of subclade i of Mitchell clade, OsAT1 is putatively functionally redundant with OsAT2, OsAT6, OsAT7, and OsAT9, since OsAT3, OsAT4, and OsAT5 are characterized as feruloyl/p-coumaroyl monolignol acyltransferases (Withers et al., 2012; Petrik et al., 2014; Smith et al., 2015a; Karlen et al., 2016; Sibout et al., 2016). OsAT9 shares the highest protein identity (62%) with OsAT10 and is expressed in most organs and development stages with the highest expression level compared to other AT members in rice (Lin et al. 2016). In support of this, a recent study found that suppression of the ortholog of *OsAT9* in *Setaria* (*SvBAHD01*) resulted in a ~60% decrease in AX feruloylation in stems, an increase in biomass saccharification, and no effect on biomass yield (de Souza et al., 2018).

To probe the function of OsAT9 in rice, we generated transgenic lines that overexpress or partially silence *OsAT9* under the control of ZmUbi promoter. Our biochemical analysis found *OsAT9* mutants significantly altered the FA:*p*CA ratio, even though the absolute changes in FA and *p*CA are relatively small. Importantly, *OsAT9* overexpression altered xylan properties and reduced the digestibility of leaves and stems.

4.3 Results

4.3.1 Overexpression of OsAT9 *increases the FA:*p*CA ratio in rice cell wall*

To test the hypothesis that OsAT9 plays a role in rice cell wall feruloylation, we overexpressed this gene using the strong constitutive maize ubiquitin promoter (*Ubi*$_{pro}$). With verified transgenic lines (*Ubi*$_{pro}$-*OsAT9*), our quantitative reverse transcriptase PCR (qRT-PCR) showed that the transcript abundance of *OsAT9* in T_1 expanding leaf was increased by 150-fold compared to the negative segregant, wild type plants (**Figure 4.1A**). Then, we measured the wall-HCA in expanding leaves at the vegetative development stage in T1 and T2 generations of transgenic line 5. To do so, we extracted the TFA-fractionated HCA from dsAIR cell wall materials by following the protocol that has been successfully used to characterize *OsAT10* and *OsAT5* (Bartley et al., 2013; Karlen et al., 2016). Our results showed that the *Ubi*$_{pro}$-*OsAT9* line possessed marginally increased hemicellulose-associated FA and decreased hemicellulose-associated *p*CA, which together leads to a significant increase in the FA:*p*CA ratio (**Figure 4.1B, C, D**). However, lignin-associated FA and *p*CA did not change in the vegetative leaf. Although hemicellulose associated ferulate is not significantly changed by *OsAT9* overexpression, the increased FA:*p*CA ratio is indicative of the overall change in acylation pattern and is consistent with OsAT9 mediating FA incorporation into rice hemicellulose.

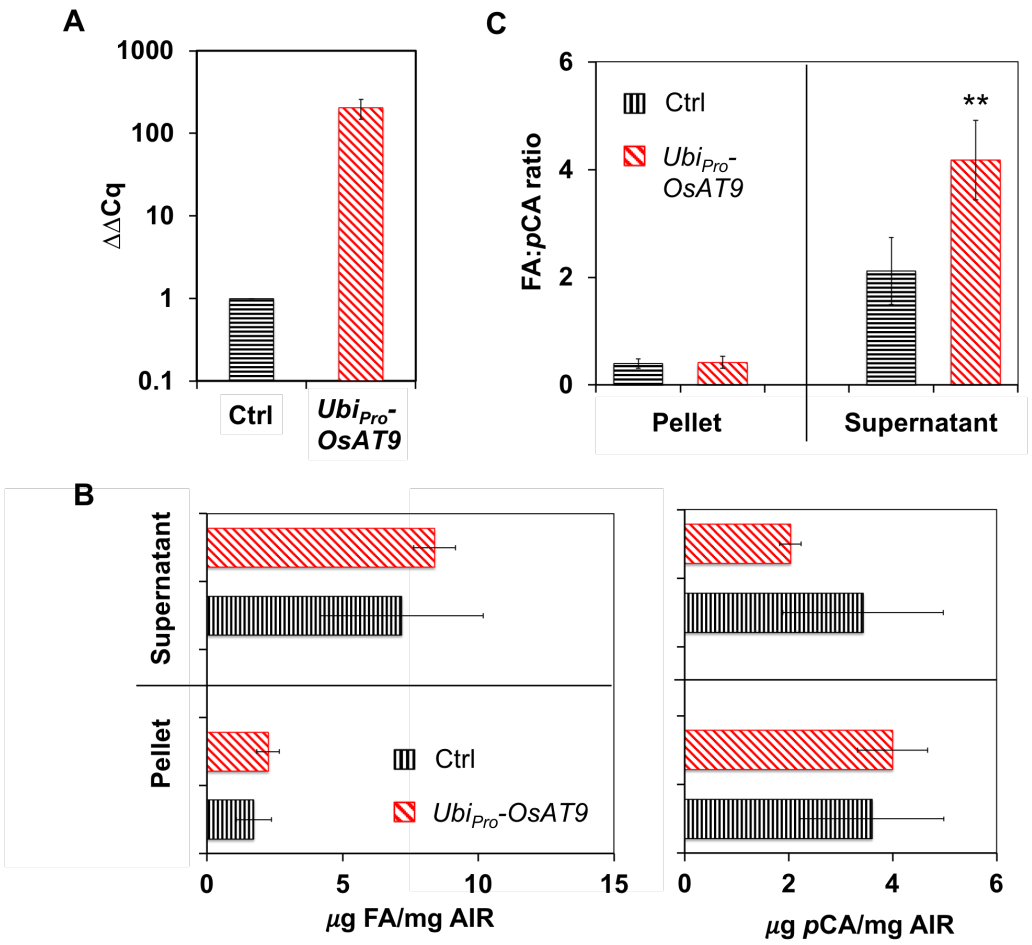


Figure 4.1 Effect of *OsAT9* overexpression on TFA-fractionated cell wall hydroxycinnamic acids.
(A) Quantitative reverse transcriptase PCR verified *OsAT9* overexpression in transgenic *Ubi*$_{pro}$-*OsAT9* lines. FA content **(B)**, *p*-CA content **(C)**, and the ratio of FA to *p*-CA **(D)** were measured in pellet (TFA-insoluble cell wall polymers, majorly lignin and cellulose) and supernatant (TFA-solubilized hemicellulose fraction). Expanding leaf at V5 from T1 transgenic plants and wild-type negative segregants without *UbiAT9* insert (Ctrl) was used for measurements. Error bars represent 2×SE of three biological replicates (**P < 0.01, by Student's t-test)

4.3.2 OsAT9 *overexpression increases xylan extractability*

Since *OsAT9* overexpression changed the content of hydroxycinnamates on arabinoxylan, we hypothesized that this would also alter xylan synthesis or deposition in cell wall. To test that, we performed immunoprofiling of extracts from wild type and transgenic cell walls with antibodies against xylan epitopes. The monoclonal antibodies used in this

study were made against the following xylan antigens from corn stove and oat. Prior to ELISA measurements, xylan was extracted sequentially with 50 mM sodium carbonate and 1 M potassium hydroxide (KOH), which solubilizes most pectin and part of xylan from cell wall polymers (Pattathil et al., 2012; DeMartini et al., 2013). For all three monoclonal antibodies used, immunoprofiling signals from extracts of overexpressed lines are increased significantly (by 50% with antibody CCRC-M138, 56% with CCRC-M144, and 52% with CCRC-M145) (**Figure 4.2A**). The literature reports that rice xylan can be substituted with glucuronic acid, arabinose, xylosy l-arabinoxyl, and galactosyl-xylosyl-arabinoxyl (Saulnier et al., 1995; Wende and Fry, 1997; Chiniquy et al., 2012). Since the antigen from corn xylan used for antibody generation is 100% xylose (Moller et al., 2008), these results imply that the abundance of non-substituted xylan may be increased or that this type of xylan is more easily released by 1M KOH, as a result of *UbiAT9* overexpression. To test the former hypothesis, we measured the total xylan and found that *OsAT9* overexpression did not change the total amount of xylose and arabinose in straw (**Figure 4.2B**), indicating that the total amount of xylan did not change. Therefore,

the increased readout of xylan detection by ELISA assays is more likely due to the improved extractability of xylan by alkaline.

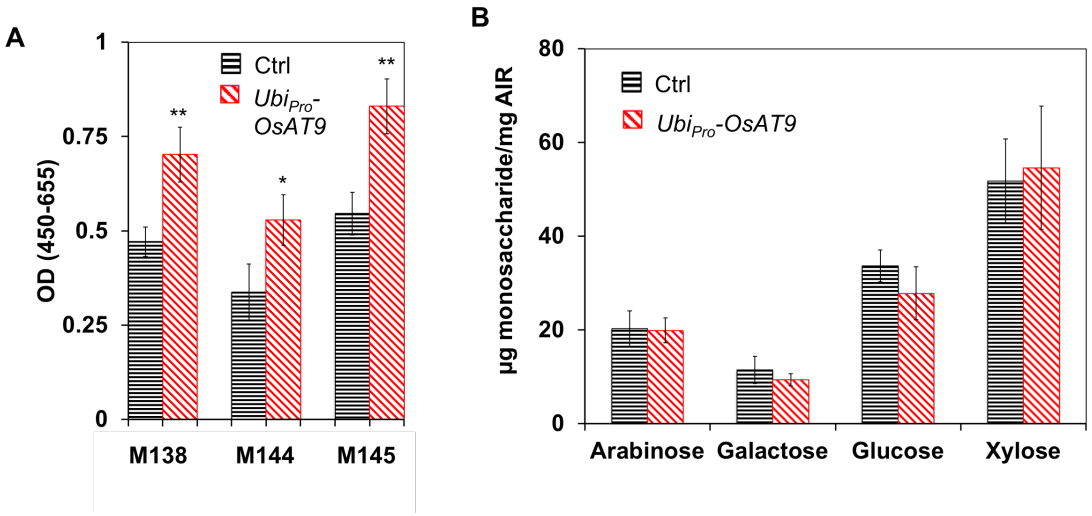


Figure 4.2 Cell wall hemicellulose profiling of *Ubi$_{pro}$-OsAT9* overexpression rice. **(A)** Enzyme immunoassays with xylan epitope (CCRC-M138, CCRC-M144, CCRC-M145 from left to right) show a higher signal in 1M KOH extracts from straw of *Ubi$_{pro}$-AT9* Line5 T1 (red) than in WT near isogenic plants (black). M138 represents Oat xylan:BSA antigen CCRC-M138; M144 and M145 represent two different antigens of Corn Stover Xylan:BSA, CCRC-M144 and CCRC-M145. Data are mean ± 2xSE of three technical replicates (*$P < 0.05$, **$P < 0.01$, by Student's t test). **(B)** Monosaccharide composition of AIR after TFA hydrolysis, separated and quantified on the HPAEC. Data are mean ± 2xSE of three biological replicates.

4.3.3 Overexpression of OsAT9 reduced enzymatic digestibility

To test whether the change of ferulates on xylan affects biomass digestibility, we examined enzyme deconstruction for mature straw, stem, leaf, and leaf sheath of *Ubi$_{pro}$-OsAT9* T2 homozygotes. All biomass used here was pretreated by mild base (6.25 mM NaOH) prior to enzyme assays. We found that overexpression of OsAT9 reduced the digestibility of mild base-pretreated stem and leaf by 26% and 19% respectively, compared to their wildtype counterparts (**Figure 4.3**). However, no significant change in degradability was observed in leaf sheath and straw. This leaf and stem-specific impact

could be explained by the variation of xylan structure and feruloylation pattern in different rice tissues. These results are consistent with previous studies on BAHD001 in seteria and Brachypodium, where its knock-down improved the digestibility of leaf and stem.

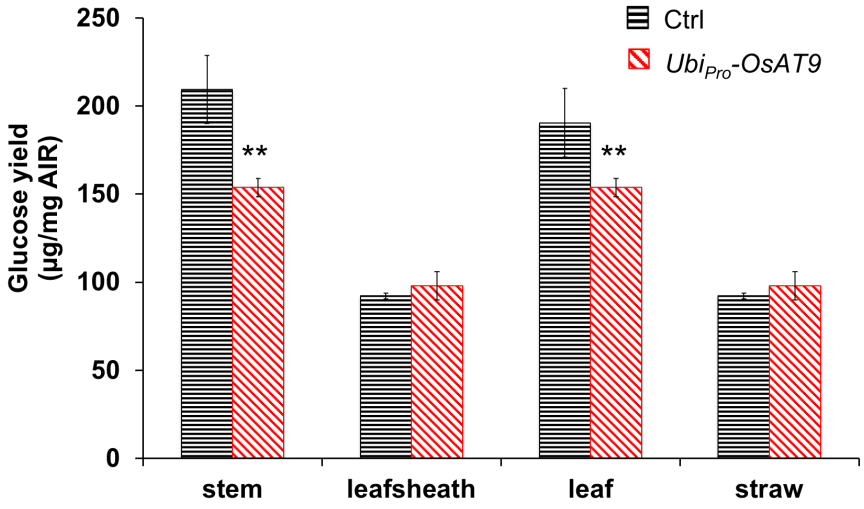


Figure 4.3 Reduced digestibility by *Ubi$_{pro}$-AT9* overexpression.
Less glucose was released from AIR of mature stem, leaf sheath, leaf, and straw of *Ubi$_{pro}$-AT9* mutant (T2 homozygotes, line 5-3) compared with wild type controls (T2 isogenic WT of *Ubi$_{pro}$-AT9* line 5-11). AIR was pretreated at 6.25 mM NaOH for 1 hr and adjusted to pH 5.0 before adding the enzyme cocktail. Error bars represent 2xSE of three biological replicates (**$P < 0.01$, by Student's t-test).

4.3.4 OsAT9 *knockdown presents a decreasing trend in the FA:*p*CA ratio in leaf*

Considering the relatively high expression level of *OsAT9* across diverse tissues, we applied RNAi technology to generate whole-plant knockdown for this gene to better understand its function. The 233-bp fragment from *OsAT9* 5'-UTR was selected to generate our RNAi hairpin construct driven by a constitutive maize ubiquitin promoter (**Figure S4.1**); this selected region is predicted to have the highest target efficiency to *OsAT9* gene rather than other genes in rice (**Text S4.1 and Table S4.2**). After transforming rice with our sequence-verified construct, we obtained 18 transgenic lines,

three of which presented successful knockdown on *OsAT9* transcripts when compared with wild type plants derived from mock transformation (**Figure S4.2A**). Phenotypically, our wild type from mock transformation were stunted and sick. Then, we measured the level of cell wall HCA in these T_0 lines and found FA reduction in several lines even with a relatively high *OsAT9* transcript level (**Figure S4.2B**). This unmatched RNA level-to-FA content observation may be due to difficulty in matching development stages between RNAi and wild-type lines. By comparing the FA content and relative gene expression level among transgenic lines, we selected line 7 with a lower gene expression and a relative reduction in FA, and line 4 with a relative higher gene expression and the most reduction in FA (**Figure S4.2**).

Consistent with *UbiAT9* overexpression, *AT9* RNAi lines only showed a significant change in the ratio of FA and *p*CA, and minimal changes in their absolute contents. At T_1 generation, we found all plants from line 7 were genetically identified as positive whereas all plants from line 4 were negative. With that, we later on used line 7 T2 as our transgenic group and line 4 T2 as wild type control such that developmentally matched tissues can be harvested for studies. We sampled matched stem and leaf tissues for qPCR validation and biochemical analysis. RNAi suppression on *OsAT9* was not significant (**Figure 4.4A**). These transgenic T2 combinatorically showed a trend of decreasing both FA and *p*CA in cell wall (**Figure 4.4B and C**), even though the difference is not

statistically significant. The FA:*p*CA ratio did not significantly change in both leaf and stem (**Figure 4.4D**).

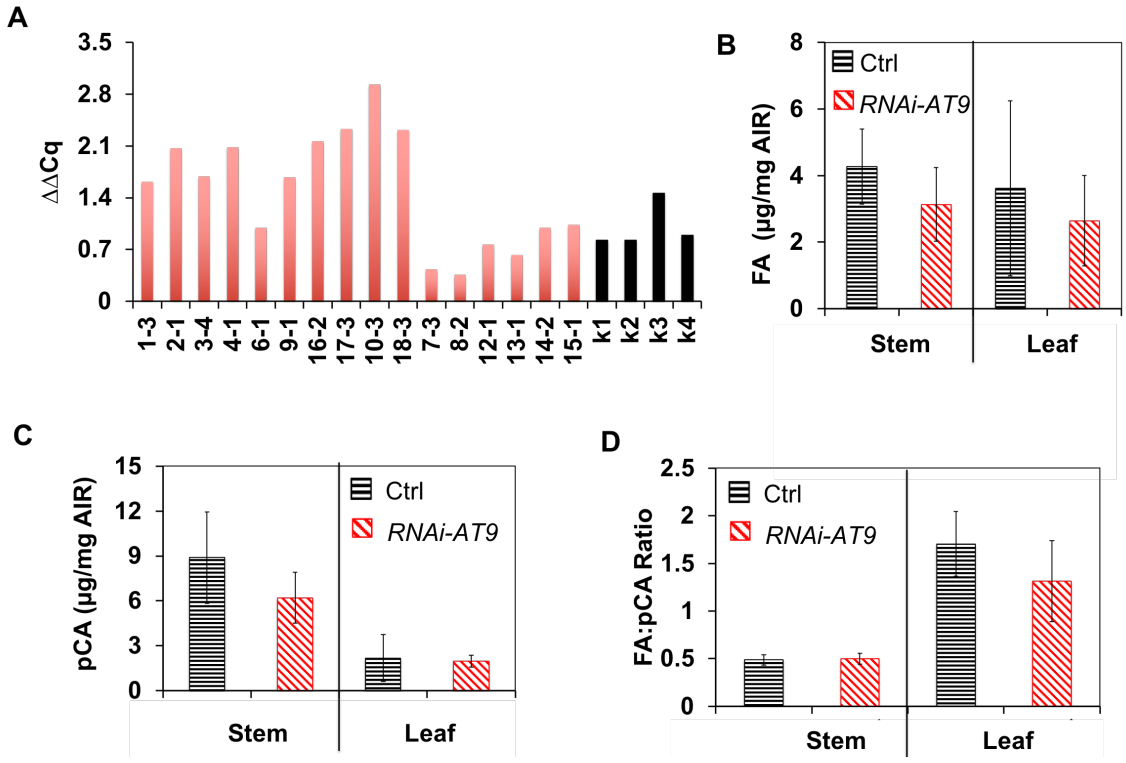


Figure 4.4 Gene expression of *RNAi-AT9* and cell wall bond HCA contents.
(A) qRT-PCR analysis of *AT9* expression in T0 generation; plant transformants (red) and negative transformants (black). **(B-D)** Measurements of FA, pCA, and the FA:*p*CA ratio in stem and leaf of *RNAi-AT9* transgenic and control lines at T1. Error bars represent 2xSE of three biological replicates.

4.4 Discussion

Ferulate esters on arabinoxylan may be one of the most important factors conferring grass cell wall recalcitrance. It is a promising strategy to improve grass digestibility by reducing the cell wall FA content to ameliorate FA-meditated crosslinking among polysaccharides and between polysaccharides and lignin. Thus, it is of significance to identify enzymes responsible for the synthesis of feruloyl arabinoxylan. In this work, we characterized a candidate feruloyl arabinoxylan synthesis gene, *OsAT9*, via constitutive overexpression

and suppression in rice. Overexpression of *OsAT9* presented a concordant impact on both FA and *p*CA esters on arabinoxylan, with an increasing and decreasing trend respectively even it not statistically significant. *OsAT9* manipulation did not change the total amount of arabinoxylan but probably changed the structure, particularly the substitution pattern of xylan. As expected, overexpression of *OsAT9* improved the enzymatic digestibility of leaf and stem in rice.

No dramatic changes in the content of HCAs

The feruloylation capacity on arabinoxylan in cell wall tissues examined here should be considered to explain why transgenic plants did not dramatically change HCAs. It is known that there are more arabinosyl substitutions on xylan (high arabi:xyl ratio) in young tissue, which will become less with maturation. Further, recent studies found feruloylation on xylan frequently occurs along with plant maturation (Lin et al., 2016). Since we only examined HCAs in vegetative young leaf, it is possible that despite more FA-arabinosyl generated by *OsAT9* overexpression at that time, the availability or activity of other proteins, such as UDP-arabinopyranose mutase (Ishii et al. 2011) and transporters responsible for xylan synthesis and decoration poses a bottleneck to accept more xylan building blocks. In another word, the high expression of OsAT9 in wild type vegetative leaf may indicate an endogenously programmed biological process able to make use of intermediates to provide sufficient products. Finding the biological bottleneck will be the key to promote the outcome of the whole cell wall biosynthetic pathway. One strategy to complement this study is to survey cell wall content of a greater diversity tissues from different development stages in the transgenic vs. wild type.

Nevertheless, knockdown of *OsAT9* might be a potential way to avoid the limitation overexpression has. However, the *RNAi-AT9* lines presented a very low suppression efficiency and did not show a significant change in wall bond HCAs. One possibility is that OsAT9 has redundant functions with other enzymes in rice.

The changed FA:*p*CA ratio indicates changes in xylan substitution pattern

Thus far, transgenic approaches to manipulate wall bond FA always correspond with an opposite alteration in xylan-associated-*p*CA substitution patterns and other alterations in xylan substitution. Several reports showed that increasing FA is accompanied with a reduction in *p*CA. Overexpression of OsAT10, a likely *p*-coumaryl-CoA arabinosyl transferase, leads to a significantly increase in *p*CA (~300%) and a reduction in FA (~40%) (Bartley et al., 2013). A concomitant change is also observed in *Setaria* where suppression of BAHD001 (the ortholog of *OsAT9*) reduced the FA but increased *p*CA (de Souza et al., 2018). In terms of effects on HCAs, the FA:*p*CA ratio is significantly increased in OsAT9 overexpression rice line. Previous studies also support that FA and *p*CA acylate the same type of xylan, since knocking out xax1 in rice generated cell wall without xylosyl-arabinosyl substitutions and meanwhile showed a reduction in both FA and *p*CA (Chiniquy et al., 2012). A possible explanation is that FA and *p*CA are competitively esterified on arabinoxylan. In our study, the changed FA:*p*CA ratio did not change the abundance of arabinoxylan but affected xylan extractability after base treatment, suggesting that the changed esterification pattern on arabinoxylan is likely due to the change in the structure, probably acylation patterns on xylan.

Thus, grasses may possess a regulatory mechanism engaging different enzymes for the synthesis of FA-arabinosyl and *p*CA-arabinosyl esters. In the cell wall co-expression network (Zhao et al., submitted), OsAT9 is correlated with OsAT2 (Os01g42870), a BAHD acyltransferase with high sequence similarity to OsAT1. It may indicate OsAT2 is responsible for acylating pCA to arabinoxylan. In comparison, OsAT9 has a relatively higher gene expression level in rice than its orthologs in *Setaria* and *Brachypodium* (BAHD001); moreover, BAHD003 did not co-express with BAHD001 in both *Setaria* and *Brachypodium*. These observations imply that the mechanism of cell wall acylation in rice is more complicated than in other grasses, and the distribution of FA-arabinoxylan or *p*CA-arabinoxylan might vary even within grass species.

Consistent effects on enzymatic digestibility

OsAT9 overexpression in rice reduced sugar release from both leaf and stem by cellulase digestion. It is consistent with the improved saccharification efficiency in RNAi-BAHD001 of *Setaria* lines. As we know, ferulate can forms cross-links and then limit biomass degradation. Even though we did not observe a significant change in wall bound FA, the increased ferulates by *OsAT9* overexpression may still undergo radical coupling by peroxidases to form dimers. Therefore, it would be worthwhile to measure and compare the abundance of dimers to further our understanding of OsAT9 function in cell wall synthesis.

4.5 Materials and Methods

4.5.1 Transgenic plants and growth condition

Rice plants were grown in a greenhouse as described before. Simply, they were grown in Turface Athletics medium/ vermiculite (1:1) mix supplemented three times per week with

fertilizer (Jack's Professional LX 15-5-15 4Ca 2Mg) at temperatures from 29° to 32°C during the day and from 24° to 25°C during the night. After germination in water, 7-day-old seedlings were transplanted to the greenhouse. Natural day lengths of less than 13 hrs were supplemented with artificial lighting.

We cloned the open reading frame of *OsAT9* (LOC_Os01g09010) from Nipponbare seedling complementary DNA (cDNA) into Gateway pDONR/Zeo with primers listed in Supplemental Table S4.1. We then recombined the cloned gene into the binary *pCAMBIA1300-Ubi-GW-Nos* construct to produce *pCAMBIA1300-UbiAT9*, which has the maize Ubi1 promoter, the 3'-terminator of nopaline synthase from *Agrobacterium tumefaciens*, and the *Hpt2* gene that confers resistance to hygromycin. The sequence-verified *pCAMBIA1300-UbiAT9* was introduced into *A. tumefaciens* strain EH105, prior to the transformation of rice cultivar Nipponbare as described (Karlen et al., 2016). Regenerated plants were grown in deep plastic containers for 1 week, transferred to pots for growth in the greenhouse under the conditions described above, and then genotyped with PCR amplification using primers of *Hpt2* and Ubi_{pro}-*OsAT9* construct (Table S4.1). We characterized the T1 progeny of Ubi_{pro}-*OsAT9* lines 5 and 7, along with negative segregant WT plants as control.

The 233 bp fragment from *OsAT9* 5'-UTR was selected as the target of RNAi (see sequence in supplement) and cloned into the vector *pHB7GW-I-WG-UBIL* to produce *RNAi-AT9*, which has hairpin designed region with cdt and pdk intron for linkage and *Hpt2* gene that confers resistance to hygromycin. The RNAi-AT9 plants were generated by Plant transformation facility of Iowa State University.

4.5.2 RT-qPCR analysis

Both *Ubi$_{pro}$-AT9* and *RNAi-AT9* were assayed using quantitative real-time reverse transcription polymerase chain reaction (**Figure 4.1**), with SYBR green qPCR supermix and primers as described previously (**Table S4.2**) (Bartley et al., 2013). Two reference genes, *Ubq5* and *Cc55*, were used in all measurements. Gene expression was compared in developmentally matched leaf tissues from vegetative developing rice plants at approximately the V5 stage. For the *Ubi$_{pro}$-AT9* line 5, T2 mutants and non-transgenic negative segregants approximately 5 weeks after sowing were harvested for qPCR analysis.

4.5.3 Cell Wall Analyses

Preparation of AIR (Alcohol insoluble residue) and dsAIR (de-starched AIR)

Leaf, leaf sheath, stem, root, and straw materials from *Ubi$_{pro}$:OsAT9* transgenic plants and isogenic wild type plants were harvested at the post-reproduction stage and dried at 45˚C for 7 days. All biomass materials were ground by a tissue homogenizer in 2-mL polypropylene tubes with one big stainless-steel ball and four small balls for 2 min at 1,400 rpm. To make AIR, ground materials were subjected to multiple washes with hot ethanol as described previously (Bartley et al., 2013) and then lyophilized afterward. The de-starched materials were obtained by treating AIR with α-amylase, amyloglucosidase and pullulanase M2 (Megazyme) (Bartley et al., 2013).

Analysis of Hydroxycinnamic Acids

HCA analysis was conducted as described previously but with minor changes (Bartley et al., 2013). AIR or dsAIR samples were saponifed with 2N NaOH for 24 hrs at room temperature and transcinnamic acid was added as the internal control. Then, the

extracts were neutralized with concentrated HCl, diluted 1:1 with pure water, and then subjected to high-performance liquid chromatography (HPLC) analysis (Bartley et al., 2013).

Hydroxycinnamate Fractionation

To run trifluoroacetic acid (TFA) fractionation assays, 2.5 mg of AIR or dsAIR samples were mixed with 500 µl of 0.05 M TFA, followed with heating at 100˚C for 4 hours. Then, supernatant and pellet were treated separately with 2M NaOH for HCA extraction. Transcinnamic acid was added as internal control for both supernatant and pellet fractions. Extracts were then subject to HPLC analysis as previously described.

Immunolocalization of glycan epitopes

We followed the published method for immunolocalization of glycan epitopes (Pattathil et al., 2012). In brief, we sequentially extracted AIR samples by using 50 mM sodium carbonate with 0.5% (w/v) sodium borohydride and 1 M potassium hydroxide (KOH) with 1.0% (w/v) sodium borohydride (pH ~10.0). Each extraction takes 24 hrs. And each extracted liquid was dialyzed 48 hrs with deionized water and then diluted to the same sugar concentration. Total sugar estimation in samples used the phenol-sulfuric acid micro plate assay. Then, we use Enzyme-Linked Immunosorbent Assay to measure the epitope abundance in samples by using glycan immunized monoclonal antibody (Moller et al., 2008).

Analysis of Monosaccharide Composition by High-Performance Anion-Exchange Chromatography (HPAEC)

To measure total xylan in this study, we measured xylose and arabinose by chromatography (Bartley et al., 2013). Simply, destarched AIR (2-5 mg) was treated with

2 M TFA at 120°C for 1 hr. Next, hydrolysate was dried using a CentriVap at 32°C. Monosaccharides were then reconstituted in pure water and analyzed by high-performance anion-exchange chromatography with pulsed amperometric detection on a Dionex ICS-3000 system equipped with an electrochemical detector and a 4- 3 250-mm CarboPac PA20 column. Monosaccharides used as external standards were purchased from Sigma-Aldrich and Alfa Aesar.

Acknowledgments

We thank Dr. Kessler, Dr. Libault, Dr. Russell, and Dr. Burgett for kind suggestions and also thank students from PBIO4313 in 2016 and 2017.

4.6 References

Akin DE, Borneman WS, Rigsby LL, Martin SA (1993) p-Coumaroyl and feruloyl arabinoxylans from plant cell walls as substrates for ruminal bacteria. Appl Environ Microbiol 59: 644–647

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, et al (2013) Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. Plant Physiol 161: 1615–1633

Buanafina MM de O, Fescemyer HW, Sharma M, Shearer EA (2015) Functional testing of a PF02458 homologue of putative rice arabinoxylan feruloyl transferase genes in Brachypodium distachyon. Planta. doi: 10.1007/s00425-015-2430-1

Casler MD, Jung HJG (2006) Relationships of fibre, lignin, and phenolics to in vitro fibre digestibility in three perennial grasses. Anim Feed Sci Technol 125: 151–161

Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD, et al (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. Proc Natl Acad Sci U S A 109: 17117–22

Damásio ARL, Braga CMP, Brenelli LB, Citadini AP, Mandelli F, Cota J, De Almeida RF, Salvador VH, Paixao DAA, Segato F, et al (2013) Biomass-to-bio-products application of feruloyl esterase from Aspergillus clavatus. Appl Microbiol Biotechnol 97: 6759–6767

DeMartini JD, Pattathil S, Miller JS, Li H, Hahn MG, Wyman CE (2013) Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. Energy Environ Sci 6: 898

Grabber JH, Hatfield RD, Ralph J (1998) Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. J Sci Food Agric 77: 193–200

Harris PJ, Hartley RD (1980) Phenolic constituents of the cell walls of monocotyledons. Biochem Syst Ecol 8: 153–160

Harris PJ, Trethewey JAK (2010) The distribution of ester-linked ferulic acid in the cell walls of angiosperms. Phytochem Rev 9: 19–33

Hatfield RD, Grabber J, Ralph J, Brei K (1999) Using the acetyl bromide assay to determine lignin concentrations in herbaceous plants: Some cautionary notes. J Agric Food Chem 47: 628–632

Iiyama K, Lam T, Stone BA (1994) Covalent Cross-Links in the Cell Wall. Plant Physiol 104: 315–320

Karlen SD, Free HCA, Padmakshan D, Smith BG, Ralph J, Harris PJ (2018) Commelinid

Monocotyledon Lignins are Acylated by p-Coumarate. Plant Physiol.

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan DDD, Helmich KE, Free HCA, Lee S, Smith BG, Lu F, et al (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci. Adv. 2:

Lam TBT, Iiyama K, Stone BA (2003) Hot alkali-labile linkages in the walls of the forage grass Phalaris aquatica and Lolium perenne and their relation to in vitro wall digestibility. Phytochemistry 64: 603–607

Lin F, Manisseri C, Fagerström A, Peck ML, Vega-Sánchez ME, Williams B, Chiniquy DM, Saha P, Pattathil S, Conlin B, et al (2016) Cell wall composition and candidate biosynthesis gene expression during rice development. Plant Cell Physiol 57: 2058–2075

Matias de Oliveira D, Finger-Teixeira A, Rodrigues Mota T, Salvador VH, Moreira-Vilar FC, Correa Molinari HB, Craig Mitchell RA, Marchiosi R, Ferrarese-Filho O, Dantas dos Santos W (2014) Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. Plant Biotechnol J n/a-n/a

Mitchell RAC, Dupree P, Shewry PR (2007) A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan. PLANT Physiol 144: 43–53

Moller I, Marcus SE, Haeger A, Verhertbruggen Y, Verhoef R, Schols H, Ulvskov P, Mikkelsen JD, Knox JP, Willats W (2008) High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles. Glycoconj J 25: 37–48

Mueller-Harvey I, Hartley RD, Harris PJ, Curzon EH (1986) Linkage of p-coumaroyl and feruloyl groups to cell-wall polysaccharides of barley straw. Carbohydr Res 148: 71–85

de O Buanafina MM (2009) Feruloylation in grasses: current and future perspectives. Mol Plant 2: 861–72

Pattathil S, Avci U, G. M (2012) Immunological approaches to plant cell wall and biomass characterization: Immunolocalization of glycan epitopes. Methods Mol Biol 908: 73–82

Petrik DL, Karlen SD, Cass CL, Padmakshan D, Lu F, Liu S, Le Bris P, Antelme S, Santoro N, Wilkerson CG, et al (2014) p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. Plant J 77: 713–26

Piston F, Uauy C, Fu L, Langston J, Labavitch J, Dubcovsky J (2010) Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. Planta 231: 677–91

Saulnier L, Vigouroux J, Thibault JF (1995) Isolation and partial characterization of

feruloylated oligosaccharides from maize bran. Carbohydr Res 272: 241–253

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61: 263–289

Sibout R, Le Bris P, Legée F, Cézard L, Renault H, Lapierre C, Legee F, Cezard L, Renault H, Lapierre C, et al (2016) Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA:monolignol transferase (PMT). Plant Physiol 170: pp.01877.2015

Smith R a, Gonzales-Vigil E, Karlen SD, Park J-Y, Lu F, Wilkerson C, Samuels a. L, Ralph J, Mansfield SD (2015a) Engineering monolignol p-coumarate conjugates into Poplar and Arabidopsis lignins. Plant Physiol 169: pp.00815.2015

Smith R a, Gonzales-Vigil E, Karlen SD, Park J-Y, Lu F, Wilkerson C, Samuels AL, Ralph J, Mansfield SD (2015b) Engineering monolignol p-coumarate conjugates into Poplar and Arabidopsis lignins. Plant Physiol 169: pp.00815.2015

de Souza WR, Martins PK, Freeman J, Pellny TK, Michaelson L V., Sampaio BL, Vinecky F, Ribeiro AP, da Cunha BADB, Kobayashi AK, et al (2018) Suppression of a single BAHD gene in *Setaria viridis* causes large, stable decreases in cell wall feruloylation and increases biomass digestibility. New Phytol. doi: 10.1111/nph.14970

Tuominen LK, Johnson VE, Tsai C-J (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. BMC Genomics 12: 236

Vogel J (2008) Unique aspects of the grass cell wall. Curr Opin Plant Biol 11: 301–7

Wende G, Fry SC (1997) 2-O-β-D-xylopyranosyl-(5-O-feruloyl)-L-arabinose, a widespread component of grass cell walls. Phytochemistry 44: 1019–1030

Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG (2012) Identification of Grass-specific Enzyme That Acylates Monolignols with p-Coumarate. J Biol Chem 287: 8347–8355

Wojtaszek P (1997) Oxidative burst: an early plant response to pathogen infection. Biochem J 322 (Pt 3: 681–92

Chapter 5 Transcriptomic dissection of tissue specificity of lignin biosynthesis- and biomass digestibility-associated genes in switchgrass

Authors: Chengcheng Zhang, Prasenjit Saha, Xiaolan Rao, Shane Cantu, Mahbobeh Lesani, Ajay Jha, Desalagn Serba, Malay Saha, Laura E. Bartley

Publication status: This chapter is in preparation as a manuscript for publication in PLoS One.

Author contributions: CZ constructed the RNA-seq pipeline, analyzed the cell wall and RNA-seq data, and wrote the manuscript.

5.1 Abstract

Switchgrass (*Panicum virgatum* L.) biomass has an important use in biofuel production; however, the recalcitrance of its biomass exerts an intrinsic hurdle for enzymatic bioconversion. With a particular interest in dissecting biomass recalcitrance-related genes, we conducted RNA sequencing of four F_1 progeny of two well-studied genotypes, AP13 and VS16, which represent the lowland and upland ecotypes, respectively. The progeny represented the digestibility extremes for the F1 population in 3 years of field studies. The greenhouse-grown tissues that we surveyed included leaf, soft stem, and hard stem and whole elongation stage 4 (E4) tillers. With the 48 samples (three replicates per sample type) we obtained RNA Seq data (>40 million paired-end reads per sample and approximately 89% unique mapping), analyzed differentially expressed genes between genotypes and between tissue types, and conducted a more detailed analysis of cell wall composition. We found that some cell wall biosynthesis genes are highly related to biomass digestibility. Further, close examination of genes encoding cell wall-decorating BAHD acyltransferases and lignin biosynthesis enzymes indicates tissue-specific expression and subfunctionalization of these families in switchgrass. This study improves our knowledge of switchgrass genes related to biomass digestibility and provides valuable information for switchgrass breeding or engineering to produce more-digestible biomass.

5.2 Introduction

The perennial grass switchgrass (*Panicum virgatum* L.) presents distinct ecotypic, phenotypic and genetic features. The U.S. Department of Energy has selected it as a model herbaceous energy crop in consideration of its high biomass yield, which is 5.2-11.1 dry tons per hectare on marginal land, and diversity of adaptation to environmental conditions (Sanderson et al., 2006). Switchgrass is generally classified into two ecotypes, upland and lowland. Uplands typically grow on droughty soils in the North; whereas lowland prefers riverine habitats and flood plains. The ecotypes exhibit distinct agronomic phenotypes. Lowlands such as the cultivar Alamo, which is represented by the genotype, AP13, is taller with a larger stem diameter, longer and wider leaf blades, fewer tillers per plant, and later flowering than upland switchgrass, such as the upland cultivar Summer, represented by the genotype, VS16 (Casler et al., 2011). Genetically, lowland cultivars are allotetraploid with 18 linkage groups distributed into two highly homologous subgenomes (2n=4x=36); upland cultivars are mostly octoploid (2n=8x=72) but sometimes tetraploid. The latter describes VS16. The allelic divergence between these ecotypes brings the chance to create F1 progeny with an enhanced heterozygosity, and a probability of inheriting superior traits from both parents, i.e., heterosis (Serba et al., 2013). Thus, hybrid breeding between the ecotypes might lead to new cultivars with combined advantageous features (Taliaferro, 2002), such as high biomass yield, broad habitat adaptability, and good biomass quality for agricultural and industrial use.

Next generation sequencing facilitates understanding of gene function in switchgrass. Earlier efforts on ESTs (Wang et al., 2012; Zhang et al., 2013) and linkage mapping of the progeny of AP13×VS16 (Serba et al., 2013) assisted the assembly of

AP13 genome. Despite technical challenges in allotetraploid genome assembly, continuous effort has been made to assemble a high-quality AP13 genome, which in the version 3.1, is approximately 1,165.7 Mb with 102,065 loci identified.

Transcriptome analysis of switchgrass has increased our understanding of this species' abiotic and biotic resistance, biomass accumulation and organ development. With the advent of illumina and SOLiD sequencing platforms, sequencing depth dramatically improved and allows for single nucleotide polymorphism (SNP) calling and differentially expressed gene (DEG) analysis. Via RNA-seq, researchers have discovered genes associated with drought stress (Meyer et al., 2014) and rust resistance in switchgrass (Serba et al., 2015). Transcriptomic profiling of different cultivars has identified genes related to distinct agronomic phenotypes. For example, comparison of transcriptomics between the early- and late-dormancy cultivars (lowland "Kanlow" vs upland "Summer") has hinted to dormancy-related mechanisms (Palmer et al., 2014); similar comparisons have provided insight into photosynthetic efficiency between ecotypes (Serba et al., 2016; Ayyappan et al., 2017), also finding DEGs between upland VS16 and lowland AP13 that are involved in biosynthesis of secondary metabolites and stress responses (Ayyappan et al., 2017). RNA-seq for development stages and organs of switchgrass, though limited, has found genes determining axillary bud initiation and development (Wang et al. 2013a), leaf development, and mineral utilization (Palmer et al. 2015).

Plant cell wall recalcitrance is a major hurdle for lignocellulose bioconversion to produce biofuels. Since hybridization can result in heterosis, we selected two major tetraploid cultivars, upland AP13 and lowland VS16, to create hybrid genotypes in F1

progeny (Serba et al., 2013); the four genotypes with most differential digestibility based on field studies were selected for further study. To explore clues on digestibility-related mechanisms, we performed deep RNA sequencing for different tissues of these genotypes. We found many genes of cell wall biosynthesis have been sub-functionalized during evolution. Greenhouse-grown clones of these genotypes however revealed only minor differences in stem and leaf tissues (Shen et al., 2009). The discovery of differentially expressed genes with tissue-specific and digestibility-associated expression features improved our understanding of switchgrass cell wall synthesis.

5.3 Materials and Method

5.3.1 Plant Materials

The switchgrass genotypes 421, 530, 514 and 541, were selected from the previously described pseudo-F1 linkage mapping population developed by crossing the lowland genotype "AP13" as the female parent with the upland genotype "VS16" as the male parent (Serba et al., 2013). They are tetraploid with an expected somatic chromosome number of 2n=4x=36. The population was constructed in 2006 at The Samuel Roberts Noble Foundation (NF), Ardmore, OK. Then, the clonal propagations of these genotypes were also grown in Red River Farm, Barneyville, OK. The phenotypic data were collected from 2008 to 2011 at Ardmore, OK and from 2009 to 2011 at Barneyville. The details of field condition and crop management practices were as described by Serba et al. 2014.

For the RNA-seq study described here, clones of the selected genotypes were grown in a greenhouse in metro-mix soil supplemented with Osmocote granular N-P-K: 14-14-14 when split and repotted, and Jack's Professional LX 15-5-15 4Ca 2Mg weekly) at temperatures from 29° to 32°C during the day and from 24° to 25°C during the night.

Natural day lengths of less than 13 hrs were supplemented with artificial lighting. We harvested plant materials from four selected genotypes at the elongation 4 stage (E4), including the whole tiller, the second leaf with sheath, and the second internode which were halved to soft and hard stems.

5.3.2 Phenotype Measurements

We measured plant height, biomass yield, and biomass properties. In vitro dry matter digestibility (IVDMD) were predicted by a Foss NIRS (Near-Infrared Spectroscopy) Systems 5000 instrument with a reflectance scanning range of 1,100-2,498 nm. Raw spectra were analyzed after calibration by VDLUFA IRS/NIT, Kassel, Germany.

Enzymatic digestion assay was conducted on the iWALL, Automated Grinding, Feeding, and Weighing System at Michigan State University, as previously described (Santoro et al. 2010). Briefly, after two different pretreatment conditions, water 90˚C 3 hrs and 1.5 mM NaOH 90˚C 3 hrs, Accellerase 1000 (Genencor, Rochester, NY) in 30 mM citrate buffer (pH 4.5) plus 0.01% sodium azide was used to hydrolyze destarched alcohol insoluble residue (dsAIR) cell wall material. Then, released glucose was assayed with the glucose oxidase/peroxidase (GOPOD) method (K-GLUC, Megazyme, Ireland).

One way ANOVA test with post-hoc Tukey HSD test were used to calculate the significance of differences in IVDMD, crude protein, NDF and enzyme digestibility.

5.3.3 RNA isolation and library construction

Switchgrass samples (described above) were ground with liquid nitrogen by mortar and pestle. Each sample type has three biological replicates. Fine powders were taken out for total RNA extraction by using RNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions. All RNAs were verified for quality and quantity by using

Agilent 2100 Bio-analyzer and Nano Drop. The cDNA libraries were constructed by using TruSeq RNA library prep kit (Illumina). In brief, RNA samples were firstly treated with DNase I to remove DNA in RNA samples. Then, messenger RNA was isolated using poly (T) oligo-attached magnetic beads and fragmentized by covaris prior to the first cDNA strand synthesis with random hexamer primers. The second cDNA strand was synthesized to create a double-stranded cDNA for end-repair, 3' adenylation, adaptor ligation, and gel purification. Finally, cDNA libraries were enriched by polymerase chain reactions and qualified by Agilent Technologies 2100 Bio-analyzer and ABI StepOnePlus Real-Time PCR System. Samples were randomly distributed to flow cell lanes.

5.3.4 RNA sequencing and expression analysis

Paired-end (2× 90 bp) sequencing reads were generated using Illumina TruSeq sequencing-by-synthesis chemistry on the HiSeq 2000 platform (Illumina). Reads were trimmed by removing adaptors, unknown nucleotides larger than 5%, and low-quality sequences (>20% of Q10 bases in one read). The trimmed reads were then aligned to *Panicum virgatum* v3.1 (Switchgrass) genome sequence (Phytozome 12) with GSNAP (Wu and Nacu, 2010), allowing 4% mismatch per read without SNP tolerance. Transcripts were quantified by HTSeq (Anders et al., 2015). Differential expression between samples was analyzed by DESeq2 in R (Love et al., 2014; Love et al., 2017) at 0.05 false discovery rate. With annotations of *Panicum virgatum* v3.1 genome, Gene ontology (GO) enrichment analysis was conducted for DEGs by GOseq in R (Young et al., 2010).

5.3.5 Identification of Mitchell clade BAHD acyltransferase gene family and lignin genes from switchgrass

To identify putative BAHD acyltransferases in switchgrass (Phytozome version 3.1), we used HMMER version 3.1 (Finn et al., 2011) with the hidden Markov model (HMM) profile of PF02458 from the Pfam database. We determined the BAHD clade of each predicted protein via comparison with the D'Auria set (D'Auria, 2006) using Clustal2 (Larkin et al., 2007) and omitted sequences that lack the region surrounding the highly conserved active-site motif, HXXXD. At the first stage, we used MEGA 5.2.2 (Tamura et al., 2011) to infer and visualize neighbor-joining phylogenetic trees for switchgrass (500 bootstraps, data not shown), from which we identified proteins most closely related to those in the Mitchell Clade. The Mitchell Clade in rice (MSU version 6.1) has been identified previously (Bartley et al., 2013). Then, the phylogenetic analysis was conducted for a maximum likelihood phylogenetic reconstruction of Mitchell Clade in rice and switchgrass, with BanAAT in BAHD Clade V from D'Auria set. We used 1000 bootstraps with the outgroup consisting of other related BAHD Clade V acyltransferases from rice, switchgrass and D'Auria set. Parameters for maximum likelihood phylogenies are as follows: amino acid substitutions according to the Jones-Taylor-Thornton model, gamma distribution of mutation rate among sites, distribution shape parameter of five, and gaps treated by partial deletion, allowing site coverage as low as 95%. The Mitchell Clade was subdivided into two subclades (i and ii) based on a relatively higher gene expression level of the subclade i in rice (Bartley et al., 2013).

For lignin pathway genes, we used both blast E-score and the function domain to determine gene families. The method is similar as reported before (Xu et al., 2009). Blast

E-score was set as 0 for searching PAL, C4H, 4CL, C3H, HCT and F5H families in switchgrass, since these families share very high conserved sequences across plant species. For CCoAoMT, COMT, CCR and CAD families, genes were identified on E-score 1E-10 and on the function domain.

5.4 Results

5.4.1 Digestibility oriented sampling design for RNA-seq

The F1 progeny of parental upland AP13 and lowland VS16 ecotypes was cultivated under two field conditions as described in the method. Cell wall and forage quality parameters were predicted from near infrared spectroscopy analysis of dried, milled biomass over four years. For this study, we selected four genotypes, hereafter named 421, 514, 530, and 541. Two of them, 421 and 530 have significantly lower NIRS-predicted In Vitro Dry Matter Digestibility (IVDMD, 35% ± 3% and 34 % ± 2% of dry matter) than the other two, 514 (40% ± 3% of dry matter) and 541 (42% ± 3% of dry matter). The former two we designated as recalcitrant genotypes (R), the latter two as digestible genotypes (D) (**Figure 1A**). In comparison, recalcitrant genotypes additionally have a lower crude protein content and higher neutral detergent fiber (NDF), which are correlated with the lower IVDMD (Mahyuddin, 2008) (**Figure S5.1 A and B**). We also found that plant height, an important agronomic feature for biomass production, is relatively higher in 421 (175 ± 10 cm) and 530 (171 ± 9 cm) when compared with 514 (148 ± 7 cm) and 541 (155 ± 9 cm) (**Figure S5.1 C**).

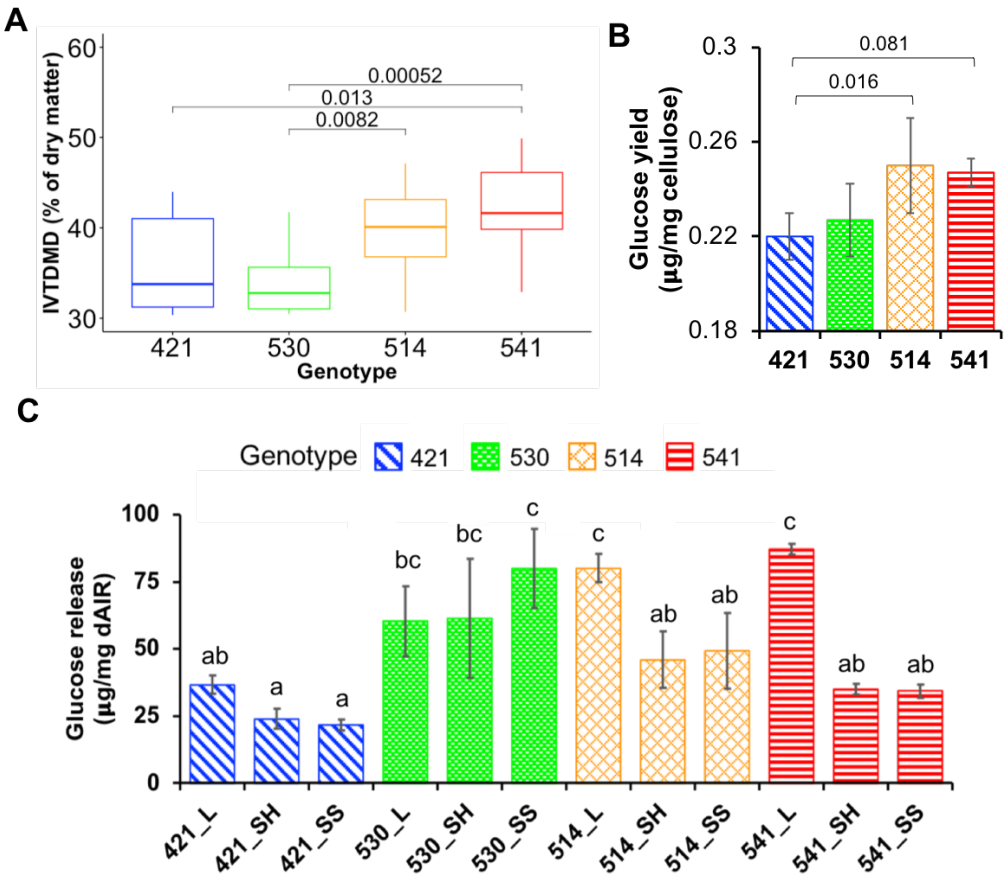
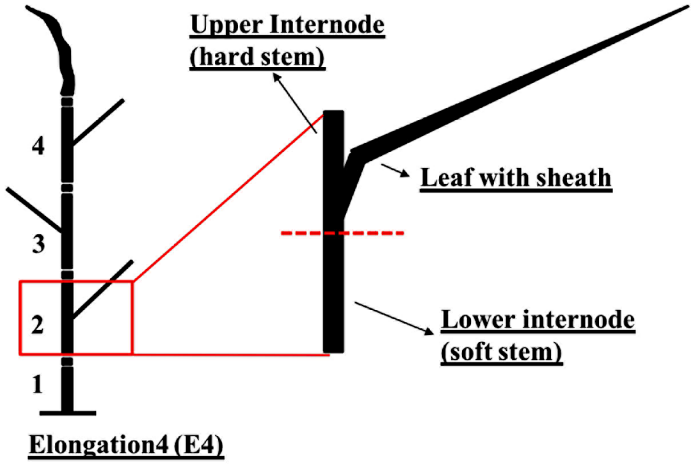


Figure 5.1 Digestibility varies among four switchgrass genotypes (421, 530, 514, and 541).
(A) IVDMD measurement of field grown materials. Significance determined by ANOVA and *p*-value were denoted. (**B)** Glucose yields from enzyme hydrolysis with pretreatment under 1.5 mM NaOH 90˚C 3 hrs for green-house grown whole tillers at E4 stage. Normalized by cellulose contents. Significance determined by ANOVA with Tukey posthoc test and *p*-value were denoted. (**C**) Glucose yields from enzyme hydrolysis for leaf (L) and stem samples (soft stem and hard stem) at E4 stage of four switchgrass genotypes (421, 530, 541, and 514) with pretreatment 1.5 mM NaOH 90˚C 3 hrs. Significance determined by ANOVA and significant groups were denoted.

To seek insight into the molecular basis of the field results, we conducted RNA Seq and chemical cell wall analysis of whole tiller (W) and dissected tissues from these four selected genotypes at the elongation stage 4 (E4) (**Figure 5.2**), including leaf with leaf sheath (L), upper half of the second internode (stem, hard; SH) and lower half of the

second internode (stem, soft; SS). We harvested four tissue samples for each genotype with three biological replicates. The overall design has some advantages. First, switchgrass at the E4 stage, which is the last vegetative stage of green-house grown switchgrass, still actively expresses diverse genes which might still effectively impact cell wall components at later development stages (Hardin et al., 2013). Second, by enzyme hydrolysis, leaf of genotypes 421, 514, and 541 yielded a higher glucose concentration than their stems (**Figure 1B**), reflecting varied digestibility between major tissues. Third, internode development-driven lignin synthesis may occur since lignin content differs in halved internodes (Sarath et al., 2007; Shen et al., 2013). Finally, in terms of digestibility of greenhouse grown biomass, the glucose yield of 421 whole tillers is 11% higher than that of 541 (**Figure 1C**), but 530 did not show significant difference to all other genotypes which is not very consistent with the observed difference from the field-grown biomass.



Group	Genotype	Sample [a]	Sample description [b]
Recalcitrant	421	421-L	Leaf with sheath
		421-SH	Hard stem (upper internode)
		421-SS	Soft stem (lower internode)
		421-W	Whole tiller
	530	530-L	Leaf with sheath
		530-SH	Hard stem (upper internode)
		530-SS	Soft stem (lower internode)
		530-W	Whole tiller
Digestible	514	514-L	Leaf with sheath
		514-SH	Hard stem (upper internode)
		514-SS	Soft stem (lower internode)
		514-W	Whole tiller
	541	541-L	Leaf with sheath
		541-SH	Hard stem (upper internode)
		541-SS	Soft stem (lower internode)
		541-W	Whole tiller

[a] Every sample has three biological replicates such that 48 samples were taken for RNA sequencing;
[b] Sample tissues are defined in the above scheme.

Figure 5.2 Diagram for sampling and summary of samples for RNA sequencing.

5.4.2 Statistics and mapping quality of 48 transcriptomes

We obtained high-quality transcriptomes from all 48 samples. An average of 43 million paired-end reads was generated for each sample, with a minimum of 35 million (530_SS2) and a maximum of 50 million (541_L2) **(Table S5.1).** After quality trimming, ~97.7% of reads were qualified across samples (**Table S5.1**); ~89.0% of qualified reads were

uniquely mapped to *Panicum virgatum* v3.1 genome, with a minimum of 84.2% (514_SH3) and a maximum of 89.8% (541_L3). These data represent >26-fold coverage of the 144.6 Mb of transcripts (using the longest transcript ".1" without considering isoforms), and include 85% of predicted protein coding genes (87,025/102,065) with assigned reads, among which 44% have >10 reads (**Table S5.1**). As expected, housekeeping genes like *Ubi10*, *GAPDH*, and *Actin1*, were consistently expressed across the samples.

To see if transcriptomic profiles distinguish our sample types, we conducted a principal component analysis (PCA) for all datasets (**Figure 5.3**). Organs are separated well by the first component, which accounts for 51% of the variance in our data; the second component contributes 17% of the variance and is able to separate genotypes. The whole tiller and stem samples tend to cluster closely depending on their digestibility or which group their genotypes fall in, recalcitrant (421 and 530) or digestible (514 and 541). A large variation was found in leaf samples of recalcitrant 421 and 530; SS and SH samples were also separated but those from genotype 530. We included all transcriptomes to detect DEGs between samples, with one exception that comparison between SS and SH samples excluded genotype 530 since our preliminary analysis found its inclusion largely confounded the DEG output. In all, our transcriptomes from whole tillers and leaf of all genotypes, and 530 excluded stem samples were good for DEG analysis.

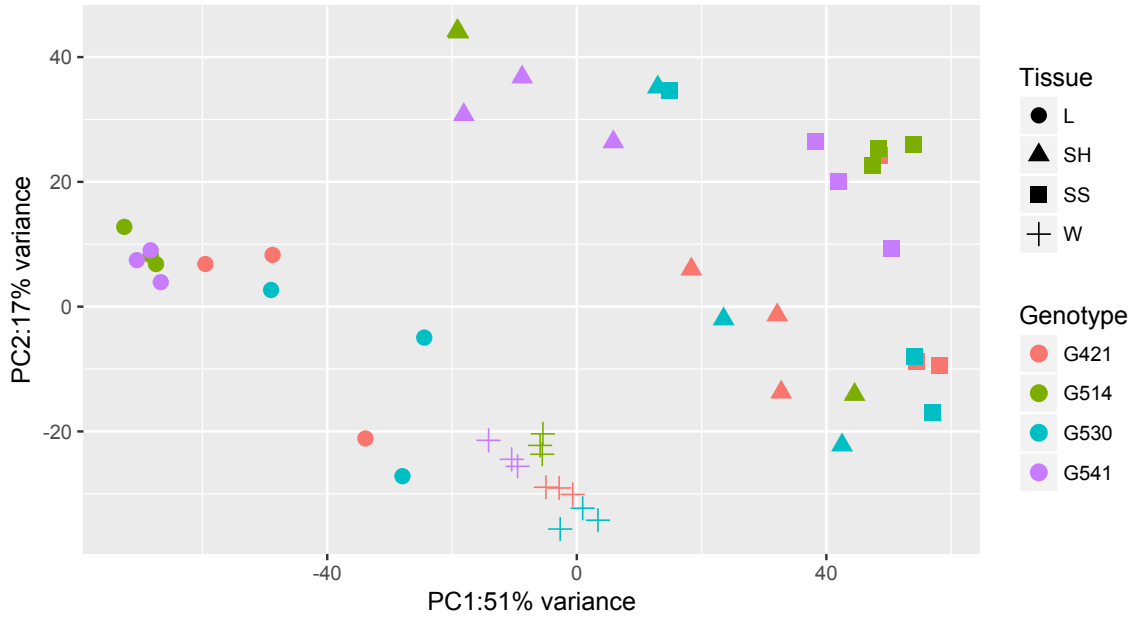


Figure 5.3 Principle component analysis of all 48 transcriptomes.
Circles denote leaf samples, triangles denote hard stems, squares denote soft stems, and crosses denotes whole tillers. Colors represent different genotypes, red for 421, green for 514, blue for 530, purple for 541.

5.4.3 Differentially expressed genes between genotypes

To explore digestibility-associated genes, we compared whole tiller transcriptomes between recalcitrant and digestible genotypes defined above. Four pair-wise comparisons we made to detect DEGs (false discovery rate, FDR < 0.01) include W421 vs W541, W421 vs W514, W530 vs W541, and W530 vs W514, where W represents whole tiller. Then, Venn diagrams identified commonly shared DEGs in recalcitrant genotypes, which include 389 up-regulated and 348 down-regulated genes, hereafter termed the WR vs WD dataset representing recalcitrant genotype-specific DEGs in whole tillers.

We then performed gene ontology (GO) enrichment analysis on both up- and down-regulated WRvsWD using go-seq (p-value < 0.05) (Young et al., 2010). The enriched GO terms with less than 10 occurrences were excluded to generate high-

confidence GO outputs. Only one GO-term, ADP-binding, belonging into the molecular function category, was overrepresented in the up-regulated subset of WRvsWD. The top 50 up-regulated genes ranked by fold change are listed in **Table S5.2**. There were no GO-terms enriched in the down-regulated subset of WRvsWD. The top 50 down-regulated genes, with several related to environment stress and ABC transporters. (**Table S5.2**).

Since the recalcitrant 421 and digestible 541 exhibited a very distinct difference in digestibility, we did similar DEG analysis and GO enrichment for their whole tiller transcriptomes (W421 versus W541). There are 2452 DEGs detected (**Figure 5.4**). In the up-regulated DEGs in 421, the most highly overrepresented biological process-related GO terms include protein folding and protein phosphorylation; the most highly overrepresented molecular function GO terms are most related to protein kinase activity (**Table 5.1**).

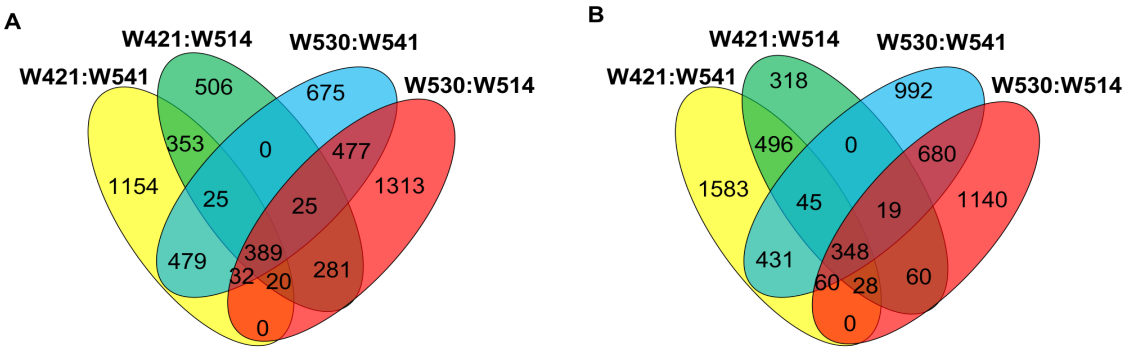


Figure 5.4 Venn diagrams show the number of DEGs identified between genotypes. Up-regulated (**A**) and down-regulated (**B**) genes in the recalcitrant group are shown here.

Table 5.1 GO-enrichment for identified DEGs.

Comparisons	GO-terms in BP [a]	GO terms in MF [b]

WR/WD		
Up-regulated	None	ADP binding
Down-regulated	None	None
W421/W541		
Up-regulated	protein phosphorylation; protein folding.	protein phosphorylation; protein kinase activity; ATP/ADP/GTP binding; nucleic acid/RNA /nucleotide binding.
Down-regulated	metabolic process; transmembrane transport.	ADP binding; transferase activity, transferring hexosyl groups.
L/S		
Up-regulated	oxidation-reduction process; metabolic process; transmembrane transport; proteolysis; cell redox homeostasis; photosynthesis; glycerol ether metabolic process; biosynthetic process; signal transduction; glycolytic process; iron-sulfur cluster assembly; isoprenoid biosynthetic process; drug transmembrane transport.	oxidoreductase activity; catalytic activity; iron ion binding; oxidoreductase activity; protein disulfide oxidoreductase activity; ATPase activity; hydrolase activity; transferase activity, transferring hexosyl groups; electron carrier activity; flavin adenine dinucleotide binding; metalloendopeptidase activity; coenzyme binding; oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; pyridoxal phosphate binding; iron-sulfur cluster binding; drug transmembrane transporter activity; antiporter activity.
Down-regulated	carbohydrate metabolic process; microtubule-based movement.	hydrolase activity, hydrolyzing O-glycosyl compounds; transferase activity, transferring acyl groups other than amino-acyl groups; microtubule motor activity; microtubule binding.
SH/SS		
Up-regulated	oxidation-reduction process; transmembrane transport; metabolic process; proteolysis; biosynthetic process; photosynthesis; cell redox homeostasis.	oxidoreductase activity; iron ion binding; catalytic activity; ATPase activity; oxidoreductase activity; protein dimerization activity; GTP binding; coenzyme binding; protein disulfide oxidoreductase activity; pyridoxal phosphate binding; methyltransferase activity.
Down-regulated	None	None

[a] Enriched GO terms ($p < 0.05$) in biological process category;
[b] Enriched GO terms ($p < 0.05$) in molecular function category.

5.4.4 Gene expression patterns in different organs

As leaf and stem are major aboveground organs expressing the majority of genes in the genome, profiling gene expression in these tissues provides a resource for our basic understanding of the switchgrass transcriptome. We defined highly expressed genes as those with >100 normalized counts. There were 19,211 highly expressed genes in leaf,, which is 31% of the total number of leaf expressed genes (61,521), and 18,777 highly expressed genes in stem, which is 31% of total expressed genes (61,020). Among these, 15,827 genes expressed in both tissues. Most highly expressed genes in leaf are predominantly involved in the synthesis of chloroplast precursors for photosynthesis and metallothionein for salinity and alkalinity stress resistance (Jin et al., 2017); in stem highly expressed genes are involved in sucrose synthesis, photosynthesis II light harvesting, the synthesis of S-adenosylmethionine, the precursor for ethylene and polyamines generation, methylation reactions (Gómez-Gómez and Carrasco, 1998), and lignin synthesis. The 20 most abundant transcripts (with >1000 normalized counts) in leaf and stem are listed in **Table S5.3** (Dominate genes in leaf and stem).

Differentially expressed genes

To identify DEGs between leaf and stem, we compared transcriptomes of leaf to soft stem (L/SS) or to hard stem (L/SH) for each genotype and then searched for common DEGs across genotypes. In total, there are 5,207 DEGs commonly shared, which is 6% of 87,025 expressed genes in both tissues and in which leaf has 2,795 genes up-regulated and 2,412 down-regulated. GO-enrichment analysis to these DEGs found terms associated with biological process, cellular component, and molecular function (**Table 5.1**). Notably, the up-regulated DEGs in leaf have highly overrepresented biological

process GO terms related to metabolic process, biosynthetic process, cell redox homeostasis, photosynthesis, proteolysis, iron-sulfur cluster assembly, glycerol ether metabolic process, drug transmembrane transport, isoprenoid biosynthetic process, oxidation-reduction process, transmembrane transport, signal transduction, and glycolytic process (**Table 5.1**). The down-regulated DEGs in leaf (i.e., up-regulated DEGs in stem) have carbohydrate metabolic process and microtubule-based movement overrepresented **(Table 5.1)**.

Tissue-specific genes

Genes that have more than 10 reads in one tissue but no more than 10 reads in other types of tissues are considered as tissue-specific genes. In this way, we identified 192 stem-specific genes and 265 leaf-specific genes (**Table S5.4**). In terms of their expression levels, 145 leaf-specific genes are highly expressed (with >100 reads) and predicted to be functional in stress responses (e.g., prolyl oligopeptidase genes, SCP-like extracellular protein, verticillium wilt disease resistance protein), flowering regulation (e.g., flowering Locus T genes), photosynthesis (e.g., heavy metal-associated domain containing protein, plastocyanin-like domain containing protein), as well as important kinase related to plant stress and development (e.g., Inositol 1,3,4-trisphosphate 5/6 kinase) (Tang et al., 2013; Marathe et al., 2018). We also identified transcription factors that are preferably expressed in leaf, including two cell wall-related R2R3-MYBs (Pavir.8KG183500 and Pavir.J611200), several NAC transcription factors involved in xylem vessel cell differentiation, and secondary cell wall (SCW) regulators (**Table 5.2**). In contrast, among stem-specific genes, there are distinct gene families with a high expression level in stem, such as auxin responsive genes, auxin and ABA crosstalk-

associated ROP-interactive CRIB motif-containing proteins, and senescence related genes (e.g., RING-H2 finger protein ATL3, leaf senescence related genes).

Table 5.2 Leaf- and stem-specific cell wall related genes.

ID	Class	Normalized counts
Leaf-specific [a]		
Pavir.8KG183500	R2R3_MYB	70.90
Pavir.J331600	NAC	46.59
Pavir.J213600	NAC	43.93
Pavir.9KG092600	NAC	34.10
Pavir.6KG122200	GT77	55.50
Pavir.5KG734000	GT77	53.01
Pavir.J660100	GT77	45.02
Pavir.5KG024000	GT37	50.22
Pavir.J660700	GT37	27.50
Pavir.2KG405200	GT2	37.07
Pavir.7KG221700	GH19	46.76
Pavir.9KG315400	GH19	16.56
Pavir.7NG218000	GH19	15.58
Pavir.3KG229100	GH1	99.46
Pavir.7NG237100	GH1	17.39
Pavir.2KG586300	F5H	177.40
Pavir.8KG172100	F5H	39.22
Pavir.9KG273100	F5H	19.65
Pavir.3NG023500	ERF	16.32
Stem-specific [b]		
Pavir.1NG495600	R2R3_MYB	186.19
Pavir.6NG055500	R2R3_MYB	135.88
Pavir.6NG055100	R2R3_MYB	122.63
Pavir.6NG153900	R2R3_MYB	39.78
Pavir.1NG188900	R2R3_MYB	29.25
Pavir.4NG062800	NAC	43.74
Pavir.J252500	MitchellClade AT10	29.58
Pavir.7NG407600	HCT	154.62
Pavir.7NG407700	HCT	34.45
Pavir.9NG603100	GT47	38.94
Pavir.6NG303700	GH19	29.86
Pavir.9NG565400	F5H	23.33

Pavir.2KG513400	COMT	44.61

[a] Leaf-specific cell wall genes refer to cell wall genes with >=10 normalized transcript counts only in leaf samples;
[b] Stem-specific cell wall genes refer to cell wall genes with >=10 normalized transcript counts only in stem samples.

Differentially expressed genes between internode segments

Internode generation in grasses provides a good model to study genes related to stem elongation development. Switchgrass internodes represent a developmental gradient with an intercalary meristem near the bottom of each internode and the most mature cells at the top of the internode (Shen et al., 2013). Thus, younger and less lignified cells in the lower part of each internode, which causes them to be soft and pliable (named soft stem, SS) compared to the upper part (named stem, hard; SH). Here, we compared transcriptomes between SH and SS sampled from the second internode of each genotype. The commonly shared DEGs accounted for 6% of 87,025 expressed genes in stem samples, consisting of 1,004 up-regulated and 215 down-regulated genes in SH. GO-enrichment analysis to these up-regulated DEGs identified overrepresented terms associated with oxidation-reduction process, transmembrane transport, metabolic process, proteolysis, biosynthetic process, cell redox homeostasis, and photosynthesis (**Table 5.1**). It is noteworthy that among the top 50 genes with the most log2 fold change in SH (**Table S5.5**), some are presumably producing cell wall-associated kinase and leucine-rich repeat protein kinase family protein, e.g. 1KG145100, Pavir.1KG287400, Pavir.7KG002000, Pavir.5KG132100; some are involved in biosynthesis of secondary metabolites, such as terpene synthase.

5.4.5 Profiling of Mitchell clade of BAHD acyltransferase gene family in switchgrass

The Mitchell clade of BAHD acyltransferase (AT) gene family plays an important role in decorating plant cell wall polymers and thereby altering the digestibility of grass biomass. Our previous study has systematically characterized the distribution of this clade in rice, Brachypodium, Soghorm, Banana, Maize, Date Palm and switchgrass genome version 1.1 (Karlen et al., 2016). In the switchgrass genome version 3.1, we found 44 members falling into the Mitchell clade. Among these 44 members, 19 belong to the subclade i where most studies have focused on thus far and in which rice only has 10 genes; the other 25 members in switchgrass are grouped to the subclade ii and this subclade usually has lower expression than the subclade i (**Figure 5.5**). It is interesting to notice that most members in the subclade i, except AT2 and AT7, have duplicated homologous genes in switchgrass but the subclade ii members are more expanded relative to rice. These duplication events are mainly due to the tetraploidy feature of switchgrass genome; whereas the singularized feature of the subclade ii members might result from transposon mediated gene transfer. Moreover, we also observed most ATs are located in the paired chromosomes (K and N), expect AT3 (Pavir.3KG136900.1 and Pavir. 3KG203500.1). Interestingly, the number of Mitchell clade genes in the subclade ii is proportional to genome size.

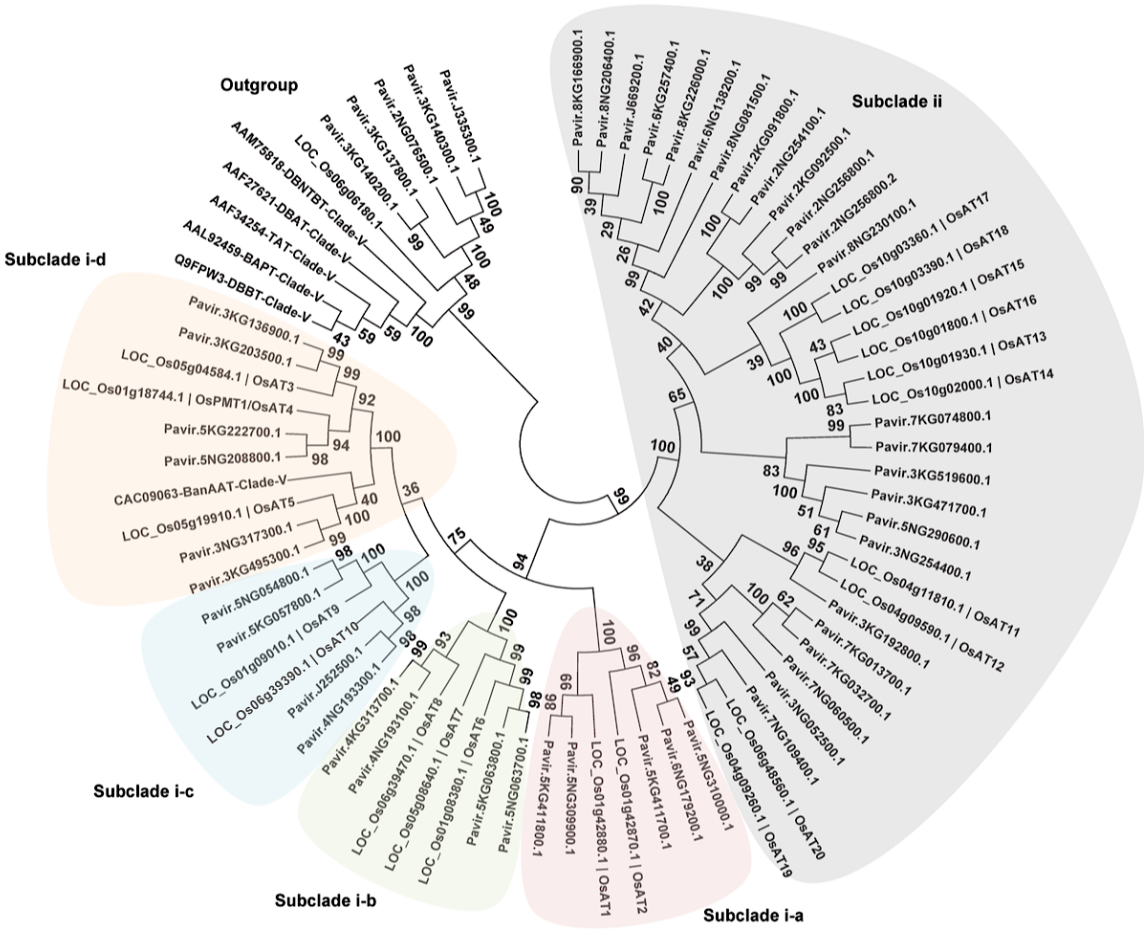


Figure 5.5 Maximum likelihood phylogeny of the BAHD acyltransferase "Mitchell Clade".
Genes from rice (MSU version 6.1) and switchgrass (Phytozome version 3.1) are included with 1000 bootstraps.

Switchgrass Mitchell ATs vary in transcript level and expression pattern across tissues (**Figure 5.6**). First, based on transcript abundance, they were grouped to high expression (AT3 > AT9 > AT1, with >1, 000 counts), moderate expression (AT2, AT6/7, and AT8, with 100-1, 000 counts), and low expression (AT4 and AT5, with < 100 counts). Second, almost half of the ATs in the subclade i, including AT1, AT3, AT6/7, AT8, and AT9, showed greater transcript abundance in stem than in leaf. Interestingly, AT6 and AT8 are likely to be stem specific genes since their expression in leaf is about 20 counts so close

to the cut-off 10 we used to define tissue-specific genes. Third, AT2 is constantly expressed across tissues; but a third AT2 homologs, Pavir.6NG179200, presented very little expression. Since the locus of the gene is not on the same chromosome pair shared by the other two homologous genes, it suggests that the genome assembly may need reorganization. Fourth, both AT4 homologs and one of AT5 homologs are almost silent in sampled tissues. The phenomenon that the other AT5 homolog has a moderate expression level may suggest sub-functionalization of homologous genes. AT10 has a little expression in all samples, which is different from the expression pattern of OsAT10 with a relatively high expression level across various tissues (Bartley et al., 2013). Fifth, all ATs in the subclade i does not show significant difference in their transcript abundance between aforementioned internode segments or genotypes. Sixth, three ATs from the subclade ii showed a higher expression level in leaf, including Pavir.2NG256800 (high in leaf but moderate in stem), Pavir.7KG013700 (leaf-specific expression), and Pavir.7NG060500 (leaf-specific expression) (**Figure S5.2).** These results will help understand functions of these AT members and potentially guide grass breeding or engineering towards the production of less recalcitrant biomass for ruminant digestion and biofuels production.

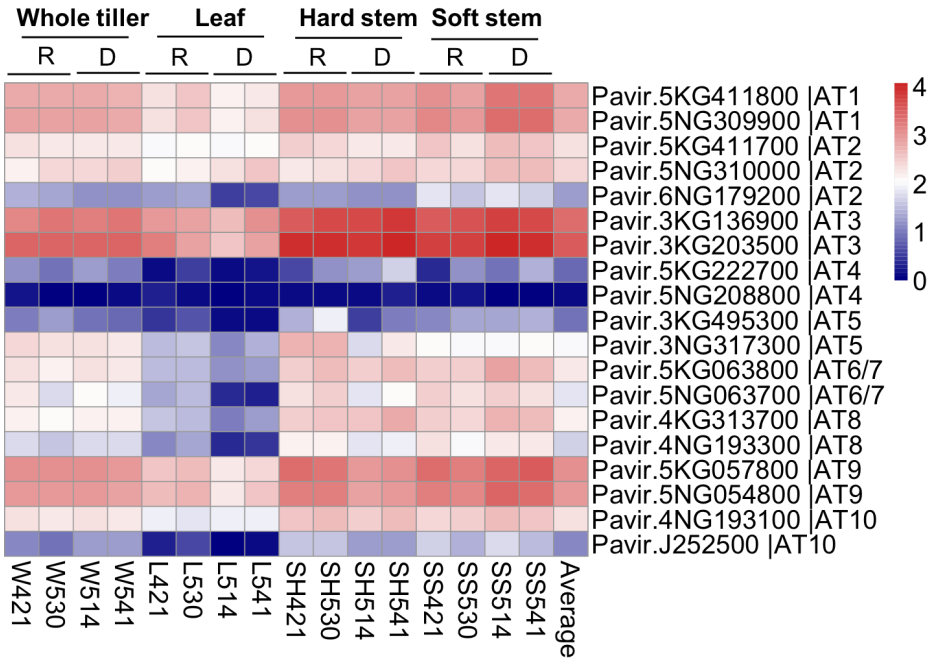


Figure 5.6 Expression pattern of Mitchell clade BAHD genes in switchgrass. From leaf to right are whole tiller (W), leaf (L), hard stem (SH), and soft stem (SS) samples from recalcitrant genotype group (R, 421 and 530) and digestible genotype group (D, 514 and 541).

5.4.6 Identification and profiling of switchgrass lignin biosynthesis genes

As lignin is the major hurdle to plant biomass digestibility and utilization, it is of great significance to understand the expression pattern of key enzymes in lignin biosynthesis. Briefly, the building blocks of grass lignin are produced by 11 enzymes, including L-phenylalanine (tyrosine) ammonia-lyase (PAL or PTAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT), 4-coumaroyl shikimate 3'-hydroxylase (C3'H), caffeoyl shikimate esterase (CSE), ferulate 5-hydroxylase (F5H), 4-coumarate CoA ligase (4CL), cinnamoyl CoA reductase (CCR), caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), caffeic acid 3-O-methyltransferase (COMT) and cinnamyl alcohol dehydrogenase (CAD) (Boerjan et al., 2003; Shen et al., 2013). Most of these enzymes have multiple isoforms that are differentially expressed

along with development and different organs and upon environmental cues (Boerjan et al., 2003). To identify these isoforms and homologous genes, we applied both HMM method and blast method to identify lignin biosynthesis genes in v3.1. There are 143 candidate lignin genes identified in transcriptomes, some of which are tissue-preferred. More specifically, we found 14 of L-phenylalanine ammonia-lyase (PAL), 3 PTALs, 7 C3'Hs, 9 C4Hs, 14 HCTs, 2 F5Hs, 2 CSEs, 11 4CLs, 26 CCRs, 22 CCoAOMTs, 5 COMTs, and 28 CADs. The expression pattern of these identified lignin genes is presented as follows (**Figure 5.7**).

PAL, the first enzyme in the biosynthesis of phenylpropanoid pathway, deaminates L-phenylalanine to t-cinnamic acid or deaminating L-tyrosine to 4-coumaric acid (4CA) in grasses (Barros et al., 2016). We identified 2 PTALs (Pavir.1KG386100 and Pavir.1NG356200) and 5 PALs (Pavir.1KG386300, Pavir.1KG386500, Pavir.1NG356400, Pavir.1NG356700, Pavir.1NG356800) in the chromosome 1, 2 PALs (Pavir.7KG237800, Pavir.7NG355500, Pavir.7NG355800) from the chromosome 7. Two P(T)ALs (Pavir.1KG386100 and Pavir.7NG355500) not only presented a higher expression level than other P(T)ALs irrespective of tissue types, but also had a stronger expression in stem than in leaf.

Subsequently, the cinnamic acid generated by PAL is hydroxylized by cinnamate-4-hydroxylase (C4H) to produce 4-coumaric acid. We found 8 C4H genes in switchgrass, which are located on three chromosomes (1, 3, and 5). C4H genes on Chr1 and Chr5 have paired homologs (Pavir.1KG264800, Pavir.1NG242900 and Pavir.5KG602000, Pavir.5NG607400). In terms of expression level, Pavir.3KG265800 is the one with the highest expression among these eight genes (>100 normalized counts); two C4Hs on the

unassembled contigs (Pavir.J661200 and Pavir.J661300) also have relatively high gene expression. C4H Pavir.5KG602000 is up-regulated in leaf compared to stem. It is noteworthy to mention that three C4Hs (Pavir.3KG265800, Pavir.5NG607400, and Pavir.J661200), presented significantly higher expression in the more digestible genotype 541.

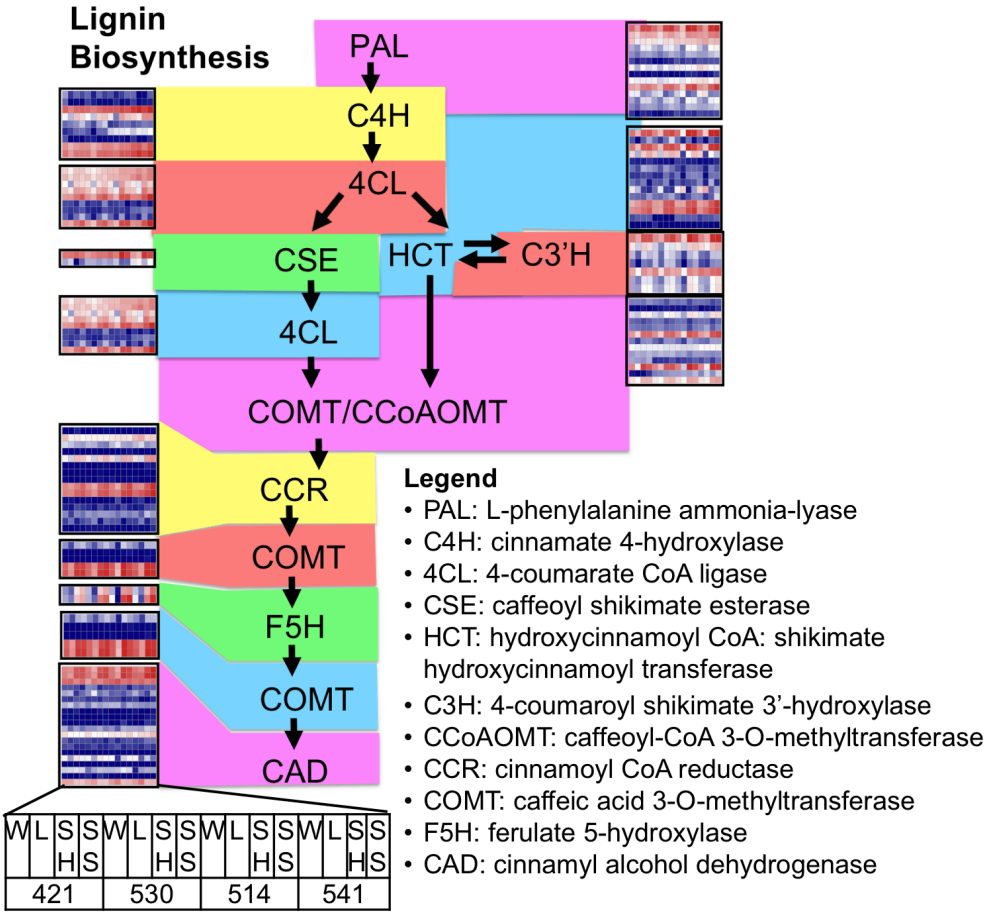


Figure 5.7 Overview of expression pattern of lignin biosynthesis genes.
Each enzyme in the pathway has at least two homologous genes, which are presented on rows of its corresponding heatmap. Columns of each heatmap indicate the same samples types from whole tiller (W), leaf (L), hard stem (SH) and soft stem (SS) from genotype 421, 530, 514, and 541 (left to right). Data are calculated by DESeq2 and represent basemean values from three biological replicates of each sample type.

4CL catalyzes the formation of activated thioesters of hydroxycinnamic acids, which may act as substrates for entry into acylation of monolignols and other branch pathways of phenylpropanoid metabolism. We identified nine 4CLs on three chromosomes. Two 4CLs with paired homologs on Chr 6 (Pavir.6KG154400 and Pavir.6NG224100) presented the highest gene expression level across tissues (>100 normalized counts). Pavir.4KG362700 and Pavir.4NG264800 prefer to express more in stem rather than leaf, whose expression patterns are consistent with the characterized switchgrass Pv4CL1 (Xu et al., 2011). The 4CL Pavir.1NG438200, has a higher expression level in the more digestible genotype 541 compared to other genotypes including 421.

We found 14 HCTs, Co-A transferases for the synthesis of shikimate or quinate, on five chromosomes (1, 2, 3, 4, and 7). The HCTs on Chr1 (Pavir.1KG378900 and Pavir.1NG344500) and Chr7 (Pavir.7KG228500 and Pavir.7NG256600) are actively expressed across most tissues with high expression level (>100 normalized counts). Pavir.3NG072400 is preferably expressed in leaf; whereas Pavir.1NG344500 and Pavir.1KG378900 prefer to be expressed in stem. There are 7 genes encoding C3H, which produces caffeoyl-shikimate/quinate ester, identified on four chromosomes (3, 5, 7, and 9). Except Pavir.9NG447400 and Pavir.5NG553300, all other five C3H genes showed preferable expression in stem, among which Pavir.3KG235800 and Pavir.5NG553300 have a relatively higher expression level. Pavir.5NG553300 also showed stronger expression in genotype 421, 530 and 514, but it showed a pretty high expression level in leaf of genotype 541.

CSEs were identified essential to switchgrass lignin biosynthesis (Ha et al., 2016).

There are two paired CSE genes identified. Although the expression of Pavir.1NG105200 is much lower than Pavir.1KG122700, both showed a trend of leaf biased expression preference in genotypes 421, 514, and 541.

Caffeoyl-shikimate/quinate is transesterified back with CoA by HCT and O-methylated in the hydroxyl group of the C3 by CCoAOMT to produce feruloyl-CoA. There are 13 CCoAOMT genes identified on chromosomes (2, 4, and 6). Pavir.6KG340200, Pavir.6NG264600 and Pavir.J369300 are expressed across all tissues we examined and showed a relative high expression level. In comparison, Pavir.4KG064500, Pavir.6KG340200, Pavir.6NG264600, and Pavir.J369300, prefer to express in stem rather than leaf. We found 2 F5H genes with paired homologs on chromosome 9. Both are preferably expressed more in stem rather than leaf.

For CCR (Escamilla-Treviño et al., 2010), we found 15 genes on chromosome 1, 2, 6, 7 and 9. Pavir.6KG279900 and Pavir.6NG147700 in pair showed the highest gene expression across tissues. Pavir.2KG274000 showed expression preference in stem. COMT catalyzes the aldehyde and alcohol precursors of S lignin by methylating 5-hydroxy-coniferaldehyde and 5-hydroxyconiferyl alcohol to yield sinapaldehyde and sinapyl alcohol. Downregulating the expression of PvCOMT reduced the cell wall recalcitrance (Fu et al., 2011a). We identified 5 genes encoding COMT on chromosome 2 and 6. Those on Chr 6 are in pair (Pavir.6KG070300 and Pavir.6NG060500) and presented the highest expression level among all COMT genes (>100 counts, in the top 20 of stem expressed genes). Even though Pavir.2KG513400 has a little expression, along with Pavir.6NG060500, they showed preferred expression in stem. For CAD (Fu et al., 2011b), there are 20 genes identified on chromosome 1, 2, 4, 6, 7, and 9.

Pavir.1KG093600, Pavir.1NG083600, and Pavir.2KG300300, have a relatively higher expression level. Pavir.6KG105700 and Pavir.9KG116700 have a moderate expression level. Pavir.1NG083600 prefers to express in stem but Pavir.9KG116700 prefers to express in leaf.

5.5 Discussion

High-quality RNA-seq

We generated over 40 million paired-end reads per sample for four genotypes and different tissues with distinct digestibility. All RNA-seq data is qualified by evaluating their mapping efficiency and bias and then allows for high-confidence transcriptomic comparisons. Due to the polyploidy nature of switchgrass, it can be challenging to differentiate reads from homologous and/or homoeologous genes. Previous studies found de novo assembly- or mapping-based analysis to RNA-seq data from seven lowland and upland ploidy cultivars gave very similar mapping rates and no substantial differences observed these two strategies, which suggests a high degree of sequence identity between homoeologous genes in switchgrass cultivars. It is also foreseeable that if the genotype of transcriptomes is not the same as the genotype of reference genome AP13, mapping and expression quantification will be inaccurate and then mislead DEG outcomes. In our work, we used a hybrid strategy to avoid this issue, which generated a very similar mapping percentage (an average of 89%) from different genotypes of F1 progeny of upland and lowland crossing (**Table S5.1**). The gene coverage reached 85%, which is consistent with a previous RNA-seq analysis of rust resistance in switchgrass (Serba et al., 2015).

Digestibility related protein

The GO-enrichment on DEGs between recalcitrant and digestible genotypes suggest a group of protein kinases or protein phosphorylation may involve in regulating biomass digestibility. Many wall associated kinases (WAKs), e.g. Pavir.5KG132100, Pavir.7KG002000, Pavir.7KG034900, and receptor like kinases (RLKs), e.g. Pavir.2KG143700, Pavir.1KG145100, Pavir.1KG287400 are actively expressed in hard stem when compared to soft stem. Protein kinase activities are overrepresented in the recalcitrant genotype 421. Kinases such as WAK and other leucine-rich repeat (LRR) RLK, play a role in regulating cell wall synthesis. Our transcriptomic analysis here may expand our knowledge on kinases regulating cell wall complexity.

Sub-functionalization of gene family members

Transcriptome analysis is a good way to assess expression divergence of duplicated genes in switchgrass. RNA sequencing has been used for this kind of study in soybean (Roulin et al. 2013). Switchgrass underwent genome duplication approxiamtely 1 million year ago (Huang et al., 2003; Yuan et al., 2015); the consequences of genome duplication are either duplicated genes which can gain new functions via neofunctionalization or functional divergence from the ancestor genes through sub-functionalization (Lynch and Conery, 2000). Some homoeologous genes as a result of sub-functionalization are silenced or activated in specific organs or development stages. By examining the expression pattern of Mitchell clade of BAHD gene family and lignin biosynthesis gene families, our knowledge of sub-functionalization of switchgrass genes is dramatically extended. We found many duplicated genes due to genome duplication, e.g. most of ATs from subclade i, but also found gene expansion probably due to transposon event, e.g. ATs from subclade ii. We do not know if they gained a different

function (neofunctionalization), but their different expression pattern may indicate subfunctionalization, e.g., *AT5*.

In general, this study expanded switchgrass transcriptome, probed biomass digestibility-related and organ-specific genes, systematically identified and profiled expression pattern of cell wall related genes. Our results not only provide evolutionary insights on gene functionalization but also bring valuable clues for cell wall studies.

Acknowledgments

Thank Dr. Tao Xu for suggestions on the pipeline construction. Thank Dr. Zhou for sharing the server for our RNA-seq analysis.

5.6 References

Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166–169

Ayyappan V, Malay Saha BC, Jyothi Thimmapuram B, Venkateswara Sripathi BR, Ketaki Bhide BP, Elizabeth Fiedler B, Hayford RK, Venu Kalavacharla B (2017) Comparative transcriptome profiling of upland (VS16) and lowland (AP13) ecotypes of switchgrass. doi: 10.1007/s00299-016-2065-0

Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. Nat Plants 2: 16050

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, et al (2013) Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. PLANT Physiol 161: 1615–1633

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54: 519–546

Casler MD, Tobias CM, Kaeppler SM, Buell CR, Wang Z-Y, Cao P, Schmutz J, Ronald P (2011) The Switchgrass Genome: Tools and Strategies. doi: 10.3835/plantgenome2011.10.0026

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 9: 331–40

Escamilla-Treviño LL, Shen H, Uppalapati SR, Ray T, Tang Y, Hernandez T, Yin Y, Xu Y, Dixon RA (2010) Switchgrass (Panicum virgatum) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. New Phytol 185: 143–155

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39: W29-37

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Chen F, Foston M, Ragauskas A, Bouton J, et al (2011a) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci 108: 3803–3808

Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon RA, Wang ZY (2011b) Downregulation of Cinnamyl Alcohol Dehydrogenase (CAD) Leads to Improved Saccharification Efficiency in Switchgrass. Bioenergy Res 4: 153–164

Gómez-Gómez L, Carrasco P (1998) Differential expression of the S-adenosyl-L-methionine synthase genes during pea development. Plant Physiol 117: 397–405

Ha CM, Escamilla-Trevino L, Yarce JCS, Kim H, Ralph J, Chen F, Dixon RA (2016) An

essential role of caffeoyl shikimate esterase in monolignol biosynthesis in Medicago truncatula. Plant J 86: 363–375

Hardin CF, Fu C, Hisano H, Xiao X, Shen H, Stewart CN, Parrott W, Dixon RA, Wang Z-Y (2013) Standardization of Switchgrass Sample Collection for Cell Wall and Biomass Trait Analysis. BioEnergy Res 6: 755–762

Huang S, Su X, Haselkorn R, Gornicki P (2003) Evolution of switchgrass (Panicum virgatum L.) based on sequences of the nuclear gene encoding plastid acetyl-CoA carboxylase. Plant Sci 164: 43–49

Jin S, Xu C, Li G, Sun D, Li Y, Wang X, Liu S (2017) Functional characterization of a type 2 metallothionein gene, SsMT2, from alkaline-tolerant Suaeda salsa. Sci Rep 7: 17914

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan DDD, Helmich KE, Free HCA, Lee S, Smith BG, Lu F, et al (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci. Adv. 2:

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–8

Love AM, Anders S, Huber W, Love MM (2017) Package ‘ DESeq2 .’

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science (80-) 290: 1151–1155

Mahyuddin P (2008) Relationship Between Chemical Component and In Vitro Digestibility of Tropical Grasses. HAYATI J Biosci 15: 85–89

Marathe A, Krishnan V, Vinutha T, Dahuja A, Jolly M, Sachdev A (2018) Exploring the role of Inositol 1,3,4-trisphosphate 5/6 kinase-2 (GmITPK2) as a dehydration and salinity stress regulator in Glycine max (L.) Merr. through heterologous expression in E. coli. Plant Physiol Biochem 123: 331–341

Meyer E, Aspinwall MJ, Lowry DB, Palacio-Mejía J, Logan TL, Fay PA, Juenger TE (2014) Integrating transcriptional, metabolomic, and physiological responses to drought stress and recovery in switchgrass (Panicum virgatum L.). BMC Genomics 15: 527

Palmer NA, Saathoff AJ, Tobias CM, Twigg P, Xia Y, Vogel KP, Madhavan S, Sattler SE, Sarath G (2014) Contrasting Metabolism in Perenniating Structures of Upland and Lowland Switchgrass Plants Late in the Growing Season. PLoS One 9: e105138

Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G (2006) Switchgrass as a

biofuels feedstock in the USA. Can J Plant Sci 86: 1315–1325

Sarath G, Baird LM, Vogel KP, Mitchell RB (2007) Internode structure and cell wall composition in maturing tillers of switchgrass (Panicum virgatum. L). Bioresour Technol 98: 2985–2992

Serba D, Wu L, Daverdin G, Bahri BA, Wang X, Kilian A, Bouton JH, Brummer EC, Saha MC, Devos KM (2013) Linkage Maps of Lowland and Upland Tetraploid Switchgrass Ecotypes. Bioenergy Res 6: 953–965

Serba DD, Uppalapati SR, Krom N, Mukherjee S, Tang Y, Mysore KS, Saha MC (2016) Transcriptome analysis in switchgrass discloses ecotype difference in photosynthetic efficiency. BMC Genomics 17: 1040

Serba DD, Uppalapati SR, Mukherjee S, Krom N, Tang Y, Mysore KS, Saha MC (2015) Transcriptome Profiling of Rust Resistance in Switchgrass Using RNA-Seq Analysis. doi: 10.3835/plantgenome2014.10.0075

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F (2009) Developmental Control of Lignification in Stems of Lowland Switchgrass Variety Alamo and the Effects on Saccharification Efficiency. BioEnergy Res 2: 233–245

Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y, Rudis MR, Tang Y, Xiao X, Jackson L, et al (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. Plant Cell 25: 4342–61

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–9

Tang Y, Tan S, Xue H (2013) Arabidopsis inositol 1,3,4-trisphosphate 5/6 kinase 2 is required for seed coat development. Acta Biochim Biophys Sin (Shanghai) 45: 549–560

Wang Y, Zeng X, Iyer NJ, Bryant DW, Mockler TC, Mahalingam R, Provart NJ (2012) Exploring the Switchgrass Transcriptome Using Second- Generation Sequencing Technology. doi: 10.1371/journal.pone.0034225

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873–881

Xu B, Escamilla-Treviño LL, Sathitsuksanoh N, Shen Z, Shen H, Percival Zhang Y-H, Dixon RA, Zhao B (2011) Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. New Phytol 192: 611–625

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, et al (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics 10: S3

Young MD, Wakefield MJ, Smyth GK (2010) goseq : Gene Ontology testing for RNA-seq datasets Reading data. Gene 1–21

Yuan S, Xu B, Zhang J, Xie Z, Cheng Q, Yang Z, Cai Q, Huang B (2015) Comprehensive analysis of CCCH-type zinc finger family genes facilitates functional gene discovery and reflects recent allopolyploidization event in tetraploid switchgrass. BMC Genomics. doi: 10.1186/s12864-015-1328-4

Zhang J-Y, Lee Y-C, Torres-Jerez I, Wang M, Yin Y, Chou W-C, He J, Shen H, Srivastava AC, Pennacchio C, et al (2013) Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). Plant J 74: 160–173

Chapter 6 Discussion and Perspective

Authors: Chengcheng Zhang, Kangmei Zhao, Laura E. Bartley

Publication status: This chapter is in preparation as a review for publication

Author contributions: CZ did the comparative analysis and wrote the paper

Lignocellulosic biofuel production is hindered by cell wall recalcitrance. Understanding characteristics and mechanisms of grass cell wall biosynthesis would be of great significance for the improvement of biomass utilization. Due to the importance of the hydroxycinnamates, ferulic acid and *p*-coumaric acid, in grass cell wall recalcitrance, we genetically characterized the function of rice BAHD acyltransferase genes that putatively decorate cell wall polymers with hydroxycinnamates, evaluated the chemical impact of these modifications on cell wall structure, and also applied transcriptomics to discover genes associated with cell wall features in the bioenergy crop switchgrass.

In Chapter 2, we reported that various angiosperm species might have convergently evolved to natively produce lignin that includes monolignol ferulate conjugates. Our measurements on these conjugates across plant species and phylogenetic analysis of the BAHD acyltransferase gene family implies that incorporation of monolignol ferulates into lignin polymers may be accomplished by a BAHD feruloyl-coenzyme A monolignol transferase, named OsAT5 in rice and its orthologs in other grasses.

In Chapter 3, we confirmed the enzyme activity of rice OsAT5 as a feruloyl monolignol transferase using heterologous expression in dicots and yeast. Interestingly, OsAT5-mediated feruloyl monolignol incorporation caused differential impacts on lignin structure and cell wall recalcitrance, depending on plant species and tissues.

In Chapter 4, overexpression of *OsAT9* in rice increased the ratio of ferulic acid and *p*-coumaric acid (FA:*p*CA) in cell wall polysaccharides, while increasing the extractability of xylan by base treatment, but reducing the enzymatic digestibility of

leaves and stems. These results indicate the important role of OsAT9 in feruloylation of rice arabinoxylan and biomass recalcitrance.

In Chapter 5, we obtained 48 high-quality transcriptomes for switchgrass samples varying in digestibility and tissue types (including leaf, soft stem, and hard stem). Differentially expressed genes between digestibility-differential genotypes and between tissues included some protein kinases and cell wall biosynthesis genes, which may function in tuning features that contribute to switchgrass digestibility. Deep transcriptomic profiling of the BAHD gene family and lignin biosynthesis genes additionally indicates sub-functionalization of these genes in switchgrass.

In all, this work significantly expanded our knowledge of plant cell wall decoration by ferulates, the role of "Mitchell-clade" BAHD members in this decoration, and their impacts on cell wall recalcitrance. This study also provides valuable information for plant breeding or engineering to produce less recalcitrant plant biomass and to understand the function of ferulate for grass biology.

Below I will project our discoveries onto the overall cell wall mechanistic studies across plant species and discuss possible functions of uncharacterized cell wall-related genes, emerging questions and hypotheses to be tested, as well as the potential industrial use of engineered plant biomass.

6.1 Gene redundancy and prediction of gene function

Gene redundancy may occur in the Mitchell Clade of the BAHD acyl-CoA-dependent acyltransferases gene family. Rice has 20 proteins in this family, falling into two subclades. In general, the subclade i members, which have been studied most, have higher expression than the subclade ii members (as is shown in **Introduction Figure 1.3**) and

are more likely to be functional in the synthesis of grass cell walls. Including the work presented in this dissertation, 4 out of 10 AT subclade i members in rice and 4 out of 12 Brachypodium ATs have been characterized. Within multi-species phylogeny, Rice ATs in the subclade i can be further subdivided to 4 small clades, with AT1 and AT2 in subclade i-a, AT6, AT7 and AT8 in subclade i-b, AT9 and AT10 in subclade i-c, and AT3, AT4 and AT5 in subclade i-d (**Figure S2.2**). Subclade i-d is well characterized as *p*-coumaroyl monolignol transferase (PMT/AT4/AT3) and feruloyl monolignol transferase (FMT/AT5), responsible for the decoration on lignin polymers (Withers et al., 2012; Petrik et al., 2014; Smith et al., 2015; Karlen et al., 2016; Sibout et al., 2016). Subclade i-c contains AT9 and AT10, which have been shown to decorate arabinoxylan by *p*-coumarate (AT10) and ferulate (AT9) (Bartley et al., 2013; de Souza et al., 2018). Subclade i-a contains the putative feruloyl monolignol transferase AT1 (Buanafina et al., 2015). Subclade i-b contains the *p*-coumaroyl monolignol transferase AT8 (Sibout et al., 2016). In terms of enzymatic activities of ATs, we speculate that subclade i-a and i-b are putatively redundant to subclade i-c and subclade i-d, respectively.

The high-quality Rice Combined mutual Ranked (RCR) gene network improves our understanding of uncharacterized members in Mitchell clade. To predict the function of uncharacterized acyltransferase, we built a one-step sub-network for each AT along with analysis of biological process GO enrichment for each sub-network; 19 out of 20 AT members are included in the RCR network. In total, we identified 279 enriched GO terms (hypergeometric p-value < 0.05); however, 240 of them (86%) are only associated with four or fewer AT sub-networks. To more confidently present possible functional

associations between ATs, we only include the GO terms shared by four or more AT sub-networks (**Figure 6.1**).

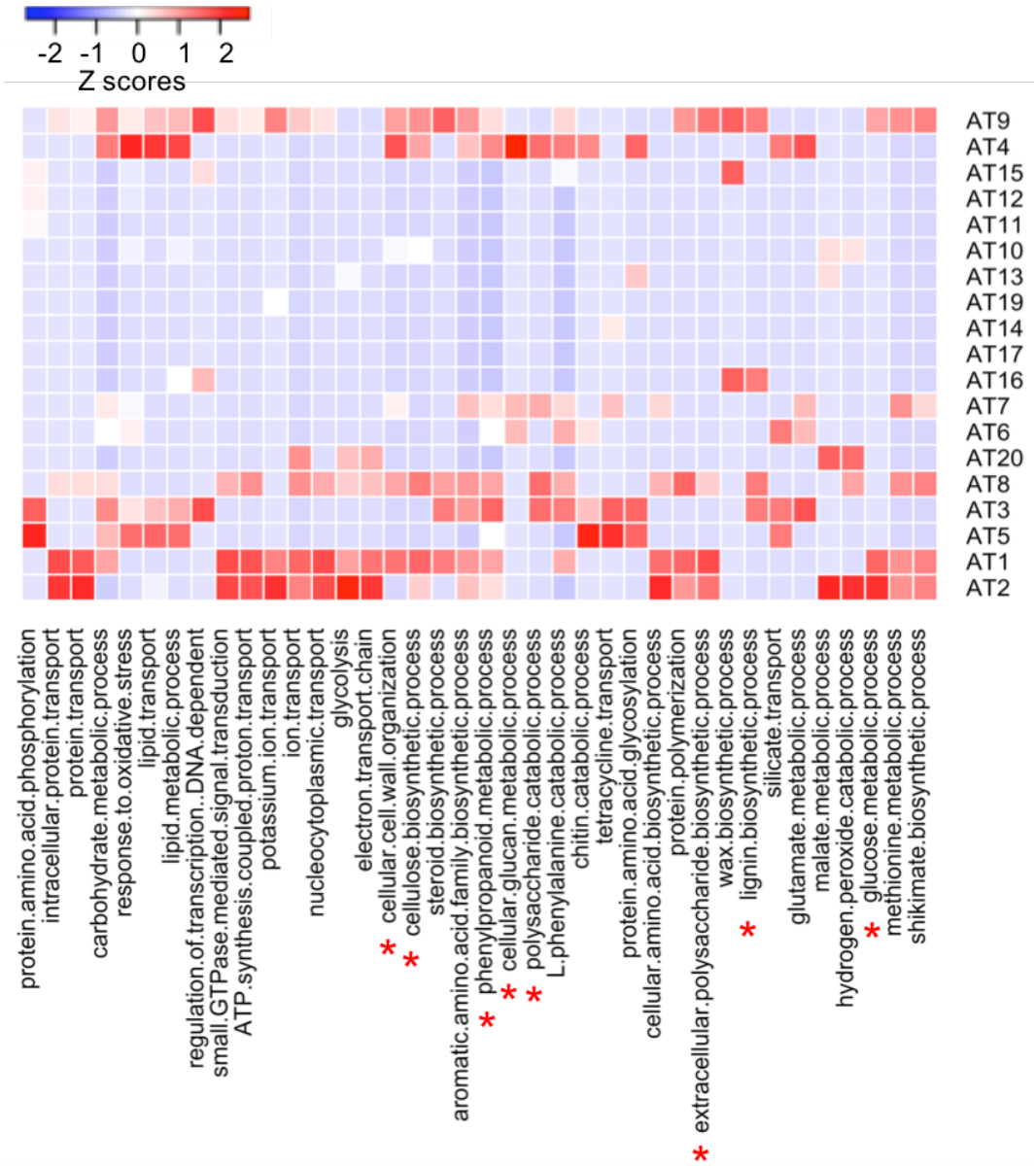


Figure 6.1 Enriched GO terms associated with components of each acyltransferase subnetwork.
One-step subnetworks were built for each AT within the RCR. Enriched GO terms in the biological process category (hyper p value < 0.05) shared by at least four AT sub-networks were included in this analysis. The heatmap shows the number of genes with selected GO terms within each AT-subnetwork normalized with z scores. Red star represents cell wall related terms.

Consistent with functional data the phenylpropanoid pathway is the most related term for Mitchell clade members. OsAT4 and OsAT5 are tightly associated with lignin biosynthesis and both of their sub-networks are enriched with phenylpropanoid pathway genes (**Figure 6.1**). The sub-networks of subclade i (AT1-AT10) are all enriched with several cell wall-related GO-terms, This meets expectations since in all known cases the acyl donor of this group of enzymes is made by the phenylpropanoid pathway. Sub-networks of AT3 and AT8 showed a close relationship with lignin biosynthesis, consistent with their role in decoration of lignin polymer. Thus, we expect that the GO term enrichment heatmap, can also provide insights on unknown acyltransferases. In this way, we speculate that OsAT1 likely relates to cellulose synthesis; AT2 may play a role in acylation of malate by hydroxycinnamates; AT6 and AT7 may be involve in cellular glucan biosynthesis.

6.2 Comparative analysis of gene expression of ATs across different species

ATs within each subclade may have a similar gene function. During evolution, some plant species may have lost or undergone an expansion of gene family members in some subclades due to genome evolution processes followed by lack of negative selection or presence of positive selection. Comparative analysis of spatial and temporal gene expression patterns of these ATs across species can imply the divergence and predominance of gene functions.

Some members in Mitchell clade i exhibit similar expression patterns across diverse grass species. Orthologs of AT1 and AT9 exist in all listed species, with a universal expression in all kinds of organs but at a lower level in leaf (**Figure 6.2**). In our genetic study on OsATs, we had trouble obtaining homozygotes of OsAT9 knock-out

mutant, indicating an indispensble role that OsAT9 is playing in plant development. Studies from my work in Chapter 4 and de Souza et al. provide genetic evidence supporting that OsAT9 may have a feruloyl-CoA transferases (FAT) activity (de Souza et al., 2018). AT1 knock-down in Brachypodium decreased feruloylation in cell wall but what components are decorated is still unknown. These results together with the fact that feruloyl-arabinoxylan is the most abundant ferulate type in grass cell walls, suggest that both AT1 and AT9 might be major FATs that decorate cell wall xylan. Due to the complexity of grass xylan (various types of side chains as shown in **Figure 1.1**), it is possible that AT1 and AT9 act on different glycan acceptors.

AT10, another member of the subclade i-c, has been genetically implicated in acting as a *p*-coumaroyl-CoA arabinosyl transferase (PAT) (Bartley et al., 2013). Overexpression of AT10 in rice and switchgrass altered both *p*CA and FA levels on arabinoxylan, without changing the substitution patterns on xylan in switchgrass, which suggest that AT10 uses *p*CA-CoA as the acyl donor instead of FA-CoA as AT9 prefers. In general, AT10 has much lower expression across species than AT9, probably in line with a lower concentration of *p*CA on xylan than that of FA (**Figure 6.2**). OsAT10 is more actively expressed in root, whereas BdAT10 and SeteriaAT10 are more in flower and orthologs in sorghum and switchgrass are even barely expressed (**Figure 6.2**). The divergent expression pattern suggests that the abundance of *p*CA-arabinoxylan may among organs and species. Thus, organ-specific examination of *p*CA-arabinoxylan in different species would be helpful to elucidate its physiological functions. Considering the possible similar function that AT1 and AT9 have, we might be able to extend our knowledge to AT2 from the subclade i-a, which also has similar expression patterns to

AT10 within each species, but at a relatively higher level. It is possible that AT2 is another *p*CA-CoA arabinoxylan transferase that is more active than AT10. Though, we also cannot exclude the possibility that AT2 uses a different glycan acceptor.

The subclade i-d has the most studied AT3 and AT4. This clade is more related to the synthesis of feruloyl/*p*-coumaroyl monolignols (Withers et al., 2012; Petrik et al., 2014; Smith et al., 2015; Karlen et al., 2016; Sibout et al., 2016). AT3 orthologs are found in all species analyzed here, which is not the case for AT4, for example, in Brachypodium and Sorghum. AT4 is more highly expressed than AT3 in rice; however, it is the opposite in other species that present a higher expression for AT3 than AT4. Brachypodium AT3 is relatively more active than that in other species (**Figure 6.2**). It is interesting that this subclade across species is usually more expressed in root, especially for AT5. This could suggest a potential physiological role that feruloyl/p-coumaroyl monolignols play in conducting water or biotic/abiotic resistance in root.

So far, we do not know much about the subclade i-b. From this subclade, BdAT8 (BdPMT2), is another PMT (Sibout et al., 2016) with a lower expression in Brachypodium; in terms of expression pattern, it is expressed biasedly towards stem in rice, but flower in Setaria, and notably, it is barely expressed in Sorghum and switchgrass (**Figure 6.2**). AT7 is actively expressed in rice rather than other species. Setaria and Sorghum even lost orthologs of AT7. In contrast, orthologs of AT6 universally exist in all species we analyzed here and present distinct expression patterns (**Figure 6.2**). AT6 is expressed more in stems of Sorghum but more in rice root. This subclade may involve in creating esters in monolignols and its potential substrate is still unclear. Our knockdown mutant of AT7 in rice decreased the feruloylation level in cell walls (Bartley

et al., 2013). A deeper characterization of this OsAT7 mutant will build knowledge of this subclade.

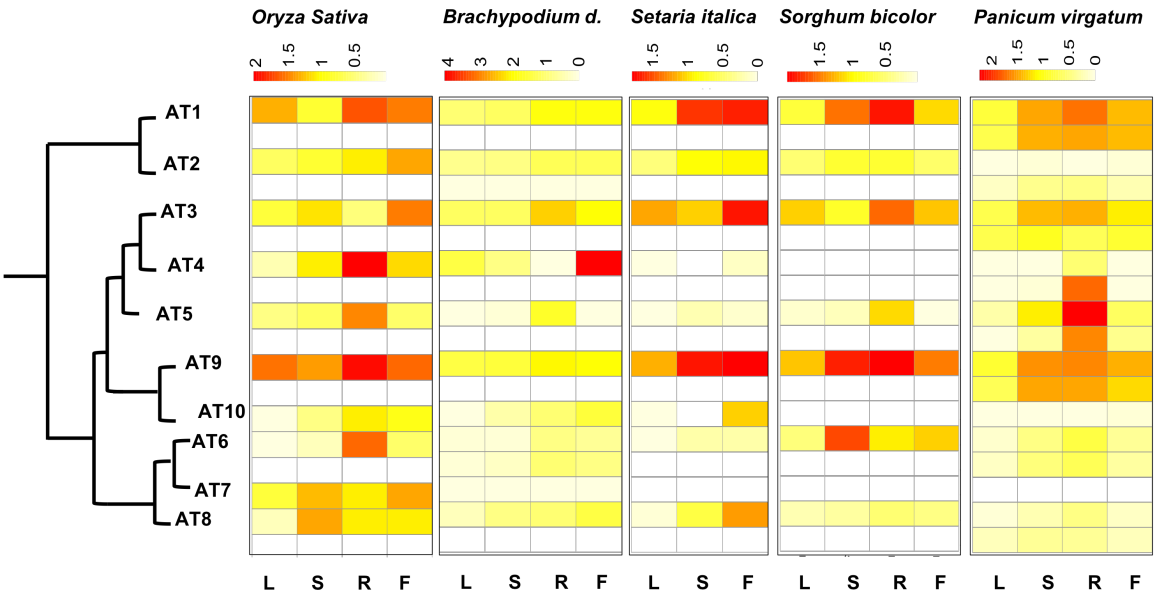


Figure 6.2 Comparative analysis of gene expression pattern of ATs across five plant species.
Oryza Sativa (rice), *Brachypodium distachyon*, *Setaria italica*, *Sorghum bicolor*, and *Panicum virgatum* are included with four types of organs, leaf (L), stem (S), root (R) and flower (F). Rice RNA-seq data are from International Rice Genome Sequencing Project in Sequence Read Archive (SRA) normalized by Rice Expression Database (RED). RNA-seq data for other species are from phytozome 12. The heatmaps were constructed based on $\log_{10}$(FPKM+1).

6.3 Physiological function of hydroxycinnamates in cell wall

When considering the engineering of plant cell wall composition to improve lignocellulosic biomass conversion, we also need to evaluate possible concomitant impacts on plant physiology, especially plant-pathogen interaction. Apart from serving as important precursors for phenolic polymer lignin, hydroxycinnamic acids also interact with root pathogens. In responding to root pathogens, many plants could release de novo synthesized HCAs into the rhizosphere and accumulate HCAs or HCA-conjugates in their

xylem sap (Mandal and Mitra, 2008; Wallis and Chen, 2012). Ferulates and diferulates also partake in defending pathogen invasion (de O Buanafina, 2009). In rice, *AT10-D1* activation has improved plant disease resistance against the most globally significant pathogen of rice, *Magnaporthe oryzae* (Mo), which causes rice blast. Improved pathogen resistance may also occur by overexpression of *OsAT5* (*AT5-D1* or Ubi_{pro}-*OsAT5* lines) but has yet to be determined. Reciprocally, inoculation of rice roots with the mycorrhizal fungus, *Rhizophagus irregularis* (Ri, formerly *Glomus intraradices*), significantly altered cell wall hydroxycinnamate content. Thus, utilizing mycorrhizal fugus will be a promising approach to study pathogen resistance of AT mutants.

6.4 Extended use of engineered lignocellulosic biomass

OsAT5 is a key target gene that can be genetically manipulated to engineer eudicot energy crops with more labile interunit linkages (i.e., esters) within lignin polymers, thereby facilitating lignin removal from polysaccharides. In this dissertation, I described differential impacts of OsAT5-produced ML-FA conjugates in dicots and grasses, showing that eudicots, such as *Arabidopsis* and Poplar trees examined in Chapter 3, have less alkaline soluble lignin compared to grasses, e.g. rice and switchgrass, which indicates a more profound role of lignin in cell wall recalcitrance in eudicots. As AsFMT expression in poplar trees introduced more ML-FAs to lignin and improved biomass digestion, it is also promising to manipulate rice OsAT5/FMT to dampen rice cell wall recalcitrance. Meanwhile, the fitness of transgenic lines should be considered since our study in the model dicots *Arabidopsis* revealed obvious perturbation of lignin biosynthesis by *OsAT5* overexpression. In terms of genetic engineering, secondary cell

wall promoters may trigger a greater effect on cell wall than constitutive promoters, which is supported by our study and others.

Increasing hydroxycinnamates in cell wall polymers, apart from reducing biomass recalcitrance, also enables lignin valorization (i.e., conversion to higher value compounds) as a promising strategy to reduce the cost of biorefining industry. Easily extractable lignin in engineered plant biomass will promote the overall extraction for biomass-derived chemical additives and possibly shape lignin valorization research and development. It is also noteworthy that some bacteria (e.g., *Amycolatopsis* sp., *Pseudomonas putida*, *Acinetobacter* ADP1, and *Rhodococcus jostii*) use hydroxycinnamic acid and other simplified lignin from plant biomass as metabolic precursors to synthesize intracellular compounds (e.g., polyhydroxyalkanoates, wax esters, triacyl glycerides) for industrial production of plastic and hydrocarbon fuels (Salvachúa et al., 2015). We envision that genetic manipulation of acyltransferases AT3, AT4, AT5, and AT10, will efficiently create less recalcitrant energy crops and lower the cost of both biofuels production and lignin valorization.

6.5 References

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sanchez ME, et al (2013) Overexpression of a BAHD Acyltransferase, OsAt10, Alters Rice Cell Wall Hydroxycinnamic Acid Content and Saccharification. PLANT Physiol 161: 1615–1633

Buanafina MM de O, Fescemyer HW, Sharma M, Shearer EA (2015) Functional testing of a PF02458 homologue of putative rice arabinoxylan feruloyl transferase genes in Brachypodium distachyon. Planta. doi: 10.1007/s00425-015-2430-1

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan DDD, Helmich KE, Free HCA, Lee S, Smith BG, Lu F, et al (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci. Adv. 2:

Mandal S, Mitra A (2008) Accumulation of cell wall-bound phenolic metabolites and their upliftment in hairy root cultures of tomato (Lycopersicon esculentum Mill.). Biotechnol Lett 30: 1253–1258

de O Buanafina MM (2009) Feruloylation in grasses: current and future perspectives. Mol Plant 2: 861–72

Petrik DL, Karlen SD, Cass CL, Padmakshan D, Lu F, Liu S, Le Bris P, Antelme S, Santoro N, Wilkerson CG, et al (2014) p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. Plant J 77: 713–26

Salvachúa D, Karp EM, Nimlos CT, Vardon DR, Beckham GT (2015) Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem 17: 4951–4967

Sibout R, Le Bris P, Legée F, Cézard L, Renault H, Lapierre C, Legee F, Cezard L, Renault H, Lapierre C, et al (2016) Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA:monolignol transferase (PMT). Plant Physiol 170: pp.01877.2015

Smith R a, Gonzales-Vigil E, Karlen SD, Park J-Y, Lu F, Wilkerson C, Samuels a. L, Ralph J, Mansfield SD (2015) Engineering monolignol p-coumarate conjugates into Poplar and Arabidopsis lignins. Plant Physiol 169: pp.00815.2015

de Souza WR, Martins PK, Freeman J, Pellny TK, Michaelson L V., Sampaio BL, Vinecky F, Ribeiro AP, da Cunha BADB, Kobayashi AK, et al (2018) Suppression of a single BAHD gene in *Setaria viridis* causes large, stable decreases in cell wall feruloylation and increases biomass digestibility. New Phytol. doi: 10.1111/nph.14970

Wallis CM, Chen J (2012) Grapevine Phenolic Compounds in Xylem Sap and Tissues Are Significantly Altered During Infection by *Xylella fastidiosa*. Phytopathology 102: 816–826

Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG (2012) Identification of Grass-specific Enzyme That Acylates Monolignols with p-Coumarate. J Biol Chem 287: 8347–8355

Appendix A: Supplementary Tables

Table S2.1. Chromatography program and MRM parameters for GC-MS/MS characterization of DFRC product mix.

Gas chromatograph	**GC-2010 Plus**
Inlet	250°C Split liner with glass wool (Shimadzu 220-90784-00) Split injection (20:1)
Column	RXi-5Sil MS 30 m x 0.25 mmx 0.25 µm (Restek 13623) Helium carrier gas Constant linear velocity 45.0 cm/sec
Oven program	150°C, hold 1 minute, ramp 10°C/minute to 300°C, hold 14.0 minutes MS interface 300°C Analysis time 30.0 minutes
Mass Spectrometer	**GCMS-TQ8030**
Ion source	250°C Electron ionization (EI) mode, 70 eV
Operation mode	Multiple Reaction Monitoring (MRM) Argon gas, 200 kPa Q1 resolution 0.8 u (Unit), Q3 resolution 0.8 u (Unit)
Detector	Electron multiplier 0.97 kV

MRM transition details

Compound name	**Retention time (min)**	**Transition 1**	**CE 1**	***Rel. Int.***	**Transition 2**	**CE 2**	***Rel. int.***	**Transition 3**	**CE 3**	***Rel. int.***	**Transition 4**	**CE 4**	***Rel. int.***
Coniferyl dihydro-*p*-coumarate diacetate (G-DH*p*CA)	*cis*: 16.99 *trans*: 17.84	370→179	10	27	370→163	14	68	370→131	22	100	370→107	30	25
Sinapyl dihydro-*p*-coumarate diacetate (S-DH*p*CA)	*cis*: 18.24 *trans*: 19.43	400→193	14	60	400→161	18	100	400→149	18	59	400→107	30	53
Coniferyl dihydroferulate diacetate (G-DHFA)	*cis*: 18.24 *trans*: 19.21	400→163	14	100	400→131	26	84	358→163	10	82	358→131	26	70
Sinapyl dihydroferulate diacetate (S-DHFA)	*cis*: 19.66 *trans*: 21.15	430→161	26	100	430→193	14	81	388→193	10	77	388→161	22	71
Diethyl 5,5'-diferulate diacetate (DEDF)*	23.30	484→442	6	100	484→396	14	19	484→350	18	73			

***DEDF was used as a control internal standard, spiked at 52.5 µg during the aqueous workup of the zinc reduction step.**

Table S2.2. Experimental results from the analysis of extract-free whole-cell-wall and enzyme lignin samples of selected eudicots.

Species	Plant material[†]	ABSL (wt%)	DFRC-released G-DHFA		DFRC-released S-DHFA		ML-FA wt% of lignin[calc.]
			mg/g ABSL	µmol/g ABSL	mg/g ABSL	µmol/g ABSL	
***Ochroma pyramidale* (WCW)**	wood[A]	18.1 ± 0.2%	0.71 ± 0.01	1.61 ± 0.02	0.10 ± 0.00	0.23 ± 0.00	4.1%
***Ochroma pyramidale* (EL)**	wood[A,EL]	–	0.66 ± 0.06	1.49 ± 0.14	0.10 ± 0.01	0.23 ± 0.02	3.8%
***Hibiscus cannabinus* (WCW)**	wood[A]	8.6 ± 0.6%	0.37 ± 0.01	0.84 ± 0.02	N/D	N/D	2.0%
***Hibiscus cannabinus* (EL)**	wood[A,EL]	–	0.47 ± 0.00	1.06 ± 0.00	0.03 ± 0.00	0.07 ± 0.00	2.6%
***Angelica sinensis* (WCW)**	root[A]	4.4 ± 0.2%	0.25 ± 0.05	0.57 ± 0.11	N/D	N/D	1.3%
***Angelica sinensis* (EL)**	root[A,EL]	–	0.26 ± 0.02	0.59 ± 0.05	N/D	N/D	1.4%

WCW = whole cell wall (samples); EL = Enzyme Lignin, i.e., Cellulysin®-digested walls in which most of the polysaccharides have been removed. Sample processing method, A, B, C or D as described in the Materials and Method section of the main paper, indicated as a superscript on the plant material description. Means and standard errors were based on duplicate technical runs. N/D signifies not detected.

Table S2.3. Experimental results from the analysis of extract-free whole-cell-wall samples of gymnosperms, magnoliids, and noncommelinid (early) monocots.

Species	Plant material	ABSL (wt%)	DFRC-released G-DHFA		DFRC-released S-DHFA		ML-FA wt% of lignin[calc.]
			mg/g ABSL	µmol/g ABSL	mg/g ABSL	µmol/g ABSL	
Cycas revoluta	petiole[A]	8.9 ± 1.4%	N/D	N/D	N/D	N/D	–
Zamia furfuracea	petiole[A]	10.5 ± 0.1%	N/D	N/D	N/D	N/D	–
Ginkgo biloba	branch[A] 5 cm[*]	16.7 ± 1.7%	N/D	N/D	N/D	N/D	–
Ephedra sp.	stem[A]	9.2 ± 1.1%	N/D	N/D	N/D	N/D	–
Welwitschia mirabilis	stem[A]	6.9 ± 1.5%	N/D	N/D	N/D	N/D	–
Gnetum gnemon	branch[A] 4 mm[*]	13.3 ± 0.9%	N/D	N/D	N/D	N/D	–
Picea sitchensis	wood[A]	18.6 ± 1.8%	N/D	N/D	N/D	N/D	–
Pinus taeda	wood[A]	18.0 ± 0.1%	N/D	N/D	N/D	N/D	–
Agathis sp.	wood[A]	48. 0 ± 0.2%	N/D	N/D	N/D	N/D	–
Dacrydium cupressinum	wood[A]	37. 0 ± 1.0%	N/D	N/D	N/D	N/D	–
Sciadopitys verticillata	branch[A] 7 mm[*]	41.0 ± 1.5%	N/D	N/D	N/D	N/D	–
Taxus brevifolia	wood[A]	31.0 ± 0.2%	N/D	N/D	N/D	N/D	–
Cupressus nootkatensis	wood[A]	18.2 ± 0.4%	N/D	N/D	N/D	N/D	–
Piper nigrum	branch[A] 7 mm[*]	9.8 ± 0.9%	N/D	N/D	N/D	N/D	–
Laurus nobilis	branch[A] 5 mm[*]	11.6 ± 0.1%	N/D	N/D	N/D	N/D	–
Magnolia acuminata	wood[A]	18.0 ±0.2%	N/D	N/D	N/D	N/D	–
Acorus calamus	leaf[A]	10.2 ± 0.5%	N/D	N/D	N/D	N/D	–
Anthurium wendlingen	petiole[A]	12.3 ± 0.9%	N/D	N/D	0.10 ± 0.01	0.21 ± 0.02	0.3%
Amorphophallus bulbiferum	leaf stem[A]	2.5 ± 0.3%	N/D	N/D	N/D	N/D	–
Pandanus tectorius	branch[A] 1 cm[*]	10.5 ± 1.5%	N/D	N/D	N/D	N/D	–
Lilium orientalis	stem[A]	22.2 ± 1.3%	N/D	N/D	N/D	N/D	–
Lilium henryi	stem[A]	15.6 ± 1.7%	N/D	N/D	N/D	N/D	–
Agave sisalana	leaf[A]	7.7 ± 1.2%	N/D	N/D	N/D	N/D	–

[*]The approximate diameter of the branch used in this study. Sample processing method, A, B, C or D as described above, indicated as a superscript on the plant material description. Means and standard errors were based on duplicate technical runs. N/D signifies not detected.

Table S2.4. Experimental results from the analysis of extract-free whole-cell-wall samples of commelinid monocots.

Species	Plant material	ABSL (wt%)	DFRC-released G-DHFA		DFRC-released S-DHFA		ML-FA wt% of lignin$^{calc.}$
			mg/g ABSL	µmol/g ABSL	mg/g ABSL	µmol/g ABSL	
Phoenix canariensis	petiole[B]	11.4 ± 0.2%	N/D	N/D	0.06 ± 0.01	0.13 ± 0.02	0.2%
Typha orientalis	leaf[B]	14.2 ± 0.5%	0.04 ± 0.01	0.08 ± 0.02	0.15 ± 0.02	0.31 ± 0.04	0.7%
Ananas comosus	core of friut[B]	5.8 ± 0.1%	0.06 ± 0.02	0.13 ± 0.04	0.88 ± 0.19	1.86 ± 0.40	3.3%
Tillandsia usneoides	whole plant[A]	9.4 ± 0.1%	0.03 ± 0.01	0.06 ± 0.02	0.17 ± 0.05	0.35 ± 0.11	0.7%
Cyperus papyrus	stem[B]	14.0 ± 0.3%	0.10 ± 0.02	0.23 ± 0.04	0.29 ± 0.06	0.62 ± 0.13	1.5%
Juncus inflexus	stem[B]	22.4 ± 0.8%	0.02 ± 0.01	0.04 ± 0.02	0.38 ± 0.01	0.81 ± 0.02	1.4%
Elegia capensis	stem[B]	14.1 ± 0.2%	0.10 ± 0.01	0.22 ± 0.02	0.54 ± 0.02	1.13 ± 0.04	2.3%
Baloskion tetraphyllum	stem[B]	16.5 ± 0.5%	0.07 ± 0.01	0.15 ± 0.02	0.73 ± 0.03	1.55 ± 0.13	2.9%
Avena sativa	stem internode[A]	22.0 ± 0.1%	0.08 ± 0.01	0.19 ± 0.02	1.29 ± 0.03	2.73 ± 0.13	4.8%
Brachypodium distachyon	stem[A]	19.1 ± 0.6%	0.02 ± 0.01	0.05 ± 0.02	0.12 ± 0.01	0.26 ± 0.02	0.5%
Miscanthus × giganteus	stem[A]	21.4 ± 0.3%	0.04 ± 0.02	0.10 ± 0.04	0.18 ± 0.02	0.39 ± 0.04	0.8%
Oryza sativa **Dongjin cultivar**	stem[D]	15.9 ± 0.9%	0.08 ± 0.03	0.18 ± 0.06	0.49 ± 0.04	1.04 ± 0.08	2.1%
Panicum virgatum	stem[A]	22.3 ± 0.6%	0.05 ± 0.01	0.12 ± 0.02	0.44 ± 0.02	0.93 ± 0.04	1.8%
Sorghum bicolor	stem[A]	17.5 ± 0.2%	0.14 ± 0.04	0.32 ± 0.09	1.70 ± 0.11	3.61 ± 0.23	6.6%
Triticum aestivum	stem internode[A]	20.7 ± 0.5%	0.13 ± 0.10	0.29 ± 0.23	0.17 ± 0.01	0.35 ± 0.02	1.2%
Zea mays **Pioneer hybrid 36H56**	stem internode[A]	17.1 ± 0.5%	0.03 ± 0.01	0.07± 0.02	0.14 ± 0.01	0.30 ± 0.02	0.6%
Dichorisandra thyrsiflora	stem[B]	12.4 ± 0.3%	N/D	N/D	1.05 ± 0.06	2.22 ± 0.13	3.6%
Musa sp.	pseudo stem[A]	11.9 ± 0.7%	N/D	N/D	0.11 ± 0.01	0.23 ± 0.02	0.4%
Strelitzia reginae	stem[B]	13.8 ± 0.5%	0.04 ± 0.01	0.08 ± 0.02	0.08 ± 0.01	0.16 ± 0.02	0.5%
Ravenala madagascariensis	petiole[A]	13.1 ± 2.3%	N/D	N/D	0.05 ± 0.01	0.10 ± 0.02	0.2%
Hedychium gardnerianum	stem[B]	10.8 ± 0.4%	N/D	N/D	0.35 ± 0.03	0.74 ± 0.06	1.2%

Sample processing method, A, B, C or D as described above, indicated as a superscript on the plant material description. Means and standard errors were based on duplicate technical runs. N/D signifies not detected.

Table S2.5. Experimental results from the analysis of extract-free whole-cell-wall samples of eudicots.

Species	Plant material†	ABSL (wt%)	DFRC-released G-DHFA		DFRC-released S-DHFA		ML-FA wt% of lignin[calc.]
			mg/g ABSL	µmol/g ABSL	mg/g ABSL	µmol/g ABSL	
Liquidambar styraciflua	wood[C]	19.8 ± 0.2%	N/D	N/D	N/D	N/D	–
Euonymus alatus	wood[A]	15.2 ± 0.7%	N/D	N/D	N/D	N/D	–
Populus deltoides	wood[C]	25.4 ± 1.5%	0.01 ± 0.00	0.03 ± 0.00	N/D	N/D	0.1%
Populus tremuloides	wood[A]	15.1 ± 0.4%	0.74 ± 0.02	1.68 ± 0.05	N/D	N/D	3.9%
Populus alba × grandidentata	wood[A]	14.8 ± 0.4%	0.29 ± 0.05	0.65 ± 0.11	N/D	N/D	1.5%
Salix babylonica	wood[C]	21.8 ± 0.9%	0.66 ± 0.09	1.49 ± 0.20	N/D	N/D	3.5%
Glycine max	stem[A]	17.7 ± 2.6%	N/D	N/D	N/D	N/D	–
Cucurbita pepo subsp. pepo	stem[A]	7.4 ± 0.2%	N/D	N/D	N/D	N/D	–
Carya ovata	wood[C]	23.2 ± 0.3%	N/D	N/D	N/D	N/D	–
Eucalyptus globulus	wood[A]	20.6 ± 0.4%	0.07 ± 0.02	0.15 ± 0.05	N/D	N/D	0.4%
Acer rubrum	wood[A]	18.0 ± 0.1%	0.22 ± 0.02	0.50 ± 0.05	N/D	N/D	1.2%
Acer saccharum	wood[A]	22.5 ± 0.5%	N/D	N/D	N/D	N/D	–
Ochroma pyramidale	wood[A]	18.1 ± 0.2%	0.71 ± 0.01	1.61 ± 0.02	0.10 ± 0.00	0.22 ± 0.00	4.1%
Hibiscus cannabinus	stem core[A]	8.6 ± 0.6%	0.37 ± 0.01	0.83 ± 0.02	N/D	N/D	2.0%
Arabidopsis thaliana	stem[A]	9.7 ± 0.8%	N/D	N/D	N/D	N/D	–
Brassica oleracea	stem[A]	17.4 ± 1.8%	N/D	N/D	N/D	N/D	–
Beta vulgaris	stem[A]	11.7 ± 0.7%	N/D	N/D	N/D	N/D	–
Bougainvillea glabra	vine stem[A]	25.0 ± 0.3%	N/D	N/D	N/D	N/D	–
Capsicum annuum	stem[A]	22.3 ± 0.1%	N/D	N/D	N/D	N/D	–
Solanum lycopersicum	stem[A]	15.0 ± 0.3%	N/D	N/D	N/D	N/D	–
Arbutus menziesii	wood[C]	17.7 ± 0.7%	N/D	N/D	N/D	N/D	–
Silphium perfoliatum	stem rind[A]	13.4 ± 0.2%	N/D	N/D	N/D	N/D	–
Angelica sinensis	root[A]	4.4 ± 0.2%	0.25 ± 0.05	0.56 ± 0.11	N/D	N/D	1.3%

Sample processing method, A, B, C or D as described above, indicated as a superscript on the plant material description. Means and standard errors were based on duplicate technical runs. N/D signifies not detected.

Table S2.6. Experimental results from the analysis of extract-free whole-cell-wall samples of plants generated in the enzyme expression study.

Species	Plant material	ABSL (wt%)	DFRC-released G-DHFA		DFRC-released S-DHFA		ML-FA wt% of lignin[calc.]
			mg/g ABSL	µmol/g ABSL	mg/g ABSL	µmol/g ABSL	
Wild-type *Brachypodium distachyon*	stem[A]	19.1 ± 0.6%	0.02 ± 0.01	0.05 ± 0.02	0.12 ± 0.01	0.26 ± 0.02	0.5%
Bdpmt-Brachypodium distachyon	stem[A]	17.8 ± 1.5%	0.05 ± 0.01	0.11 ± 0.02	0.10 ± 0.01	0.21 ± 0.02	0.6%
Wild-type *Arabidopsis thaliana*	stem[A]	9.7 ± 0.8%	N/D	N/D	N/D	N/D	–
proCESA7::OsPMT-Arabidopsis thaliana Line A	stem[A]	7.6 ± 0.8%	N/D	N/D	N/D	N/D	–
proCESA7::OsPMT-Arabidopsis thaliana Line B	stem[A]	11.2 ± 0.1%	N/D	N/D	N/D	N/D	–
proCESA7::OsPMT-Arabidopsis thaliana Line C	stem[A]	10.4 ± 0.1%	N/D	N/D	N/D	N/D	–
Wild-type *Oryza sativa*[*] Line 9 (isogenic wild type)	straw[D]	15.4 ± 0.3%	0.06 ± 0.01	0.14 ± 0.03	0.50 ± 0.11	1.05 ± 0.23	2.0%
Ubipro::OsAT5-Oryza sativa[*] Line 9	straw[D]	13.0 ± 1.1%	0.35 ± 0.09	0.78 ± 0.21	2.87 ± 0.54	6.08 ± 1.15	11.6%
Wild-type *Oryza sativa*[*] Line 10 (isogenic wild type)	straw[D]	15.4 ± 0.3%	0.05 ± 0.00	0.11 ± 0.00	0.44 ± 0.05	0.94 ± 0.10	1.8%
Ubipro::OsAT5-Oryza sativa[*] Line 10	straw[D]	14.6 ± 0.8%	0.30 ± 0.14	0.69 ± 0.31	2.50 ± 1.32	5.29 ± 2.79	10.1%
Wild-type *Oryza sativa*[*] Line 17 (isogenic wild type)	straw[D]	15.4 ± 0.3%	0.08 ± 0.00	0.18 ± 0.01	0.62 ± 0.01	1.31 ± 0.01	2.5%
Ubipro::OsAT5-Oryza sativa[*] Line 17	straw[D]	13.0 ± 1.1%	0.38 ± 0.07	0.86 ± 0.16	3.18 ± 0.43	6.74 ± 0.91	12.8%
Wild-type *Oryza sativa*[†] Line 11 (PFG_2A-20021)	straw[D]	16.1 ± 0.6%	0.08 ± 0.01	0.18 ± 0.03	0.49 ± 0.02	1.04 ± 0.05	2.1%
OsAT5-D1-Oryza sativa[†] Line 10 (PFG_2A-20021)	straw[D]	15.6 ± 0.9%	0.30 ± 0.03	0.68 ± 0.06	2.79 ± 0.14	5.91 ± 0.30	11.1%
Wild-type Poplar[‡] Line P_{39}	wood[A]	15.7 ± 1.7%	0.29 ± 0.05	0.66 ± 0.11	N/D	N/D	1.5%
CesA8::AsFMT-YFP-Poplar[‡] Line 7	wood[A]	14.8 ± 0.4%	1.64 ± 0.01	3.71 ± 0.2	2.50 ± 0.01	5.30 ± 0.02	17.2%
CesA8::AsFMT-YFP-Poplar[‡] Line 6	wood[A]	18.0%[§]	1.96[§]	4.43[§]	2.13[§]	4.51[§]	17.6%[§]
CesA8::AsFMT-YFP-Poplar[‡] Line 8	wood[A]	22.2%[§]	1.06[§]	2.40[§]	0.46[§]	0.97[§]	7.2%[§]

Cultivar / cross line indicated after species name: [†]Nipponbare cultivar. [†]Dongjin cultivar [‡]*Populus alba* × *grandidentata*. Sample processing method, A, B, C or D as described above, indicated as a superscript on the plant material description. [§]Indicated when previously reported values for released ML-DHFAs per gram Klason lignin. Means and standard errors were based on triplicate technical runs. N/D signifies not detected.

Table S2.7. Primers used in this study.

A. Primers used for genotyping

Line	5' primer name	5' primer sequence (5→3)	3' primer name	3' primer sequence (5→3)	primer pair for T-DNA::plant junction
2A-20021	L0.5	TTGGGGATCCTCTAGAGTCGAG			
2A-20021	Pam1-13F	TGACTGAAGGTCGAGAACGA	Pam1-13R_15LB	GTTACATGATGCCTTGTCAAG	Pam1-13R_15LB / L0.5
Ubi::OsAt5	Hyg-3	TCCACTATCGGCGAGTACTTCTACACA	Hy-4	CACTGGCAAACTGTGATGGACGAC	
Ubi::OsAt5	OsAT5_pri941	CTGAGGATGTTGCCACC	35S/ubiGUSnos+R	ATGACACCGCGCGCGATAATTTATCCTAG	

B. RT-qPCR primers

ID	5' primer name	5' primer sequence	3' primer name	3' primer sequence	Purpose	Reference
LOC_Os05g19910	os05g19910_543F	CATCACTGCAGTAGCTGAATTGG	os05g19910_634R	GCTTAGGTGGGTTCGGTATCAGC	experimental	Piston *et al.* 2010 (40)
LOC_Os05g19900	Os05g19900_qF1	TGCCGCACTCCGCTTTTGTGGGA	Os05g19900_qR1	TGTCGAGACTGAGCCACTCGCAATC	experimental	this study
LOC_Os05g19920	Os05g19920_qF1	CGACAACAGCGGATCAGGTGGC	Os05g19920_qR2	ACCTGATCCGCTGTTGTCGAGGT	experimental	this study
LOC_Os05g19930	Os05g19930_qF2	ACACAACCACCAGGAGCCAT	Os05g19920_qR2	AGTGACTTGGAAAATGTCACTTCTT	experimental	this study
	CC55 R1-5'	AAGGAGAAAGCCGAACAACG	CC55 RT1-3'	TCCTCAAGTTTCTTCCTGTAGGC	control	Bartley *et al.* 2013 (17)
	UBQ5 RT 1-5'	ACCACTTCGACCGCCACTACT	UBQ5 RT1-3'	ACGCCTAAGCCTGCTGGTT	control	Jain *et al.* 2006 (41)

Table S3.1. Measurement of *p*CA released by mild alkaline hydrolysis of AIR of mature inflorescence stems from Arabidopsis wild type (Col-0) and transgenic lines expressing *OsAT4* under the control of 35S promoter and the C4H promoter respectively.

Genotype	*p*CA (µg/mg)	Lignin-*p*CA (µg/mg)
Col-0 (control of 35S-AT4)	0.05±0.01	0.12±0.00
35S-AT4 (T1)	0.09±0.02	
Col-0 (control of C4Hpro-OsAT4)	0.03±0.01	0.125±0.003
C4Hpro-OsAT4 (T1)	5±1 ***	1.2±0.3 **

The data for *35S-AT4* transgenic plants represent mean values (2xSE) from 6 bio-replicates from T2. The data for *$C4H_{pro}$-OsAT4* represent mean values (2xSE) from 10 independent transgenic plants from T1.

Table S3.2. Measurement of *p*-coumaric acid (*p*CA) released by mild alkaline hydrolysis of AIR of mature inflorescence stems from Arabidopsis wild type (Col-0) and transgenic lines expressing *OsAT5* under the control of 35S promoter.

Genotype	FA (µg/mg)
Col-0	0.014 ±0.002
35S-AT5 (T1)	0.016±0.005

The data for 35S-AT5 represent mean values (and 2xSE) from 5-7 bioreplicates from T1 generation.

Table S3.3. DFRC detected *p*CA-MLs in *$C4H_{pro}$-OsAT5* transgenic lines.

Genotype	G-OH*p*CA (µg/mg ABSL)	S-OH*p*CA (µg/mg)
Col-0	0	0
$C4H_{pro}$-OsAT5 Line1	0.027±0.001	0.54±0.02
$C4H_{pro}$-OsAT5 Line5	0.017± 0.006	0.62±0.14
$C4H_{pro}$-OsAT5 Line7	0.019±0.001	0.63±0.02

The data for *$C4H_{pro}$-OsAT5* represents mean values (and 2xSE) from pooled T2 generation from 40 plants for each line and two technical replicates.

Table S4.1 BlastN results for RNAi selected sequence**.**

Locus	Description	Hit Score	E-value	Num HSPs	Top Query Cov	Top Id
LOC_Os01g09010	genomic|transferase family protein	1165	1.4e-46	1	100.00%	100.00%
LOC_Os08g12020	genomic|retrotransposon protein, Ty3-gypsy subclass	137	0.096	2	45.06%	63.30%
LOC_Os10g08240	genomic|retrotransposon, centromere-specific	175	0.11	1	94.85%	59.47%
LOC_Os05g23180	genomic|transposon protein, unclassified	167	0.22	1	52.36%	63.71%
LOC_Os12g28500	genomic|retrotransposon protein, unclassified	158	0.46	1	52.36%	62.90%
LOC_Os05g19900	genomic|retrotransposon protein, Ty3-gypsy subclass	159	0.47	1	52.36%	64.29%
LOC_Os04g06144	genomic|retrotransposon protein, Ty3-gypsy subclass	157	0.55	1	46.35%	64.55%
LOC_Os01g47970	genomic|retrotransposon protein, unclassified	156	0.55	1	34.33%	71.26%
LOC_Os07g20720	genomic|retrotransposon protein, Ty3-gypsy subclass	155	0.63	1	50.64%	66.94%
LOC_Os06g25230	genomic|retrotransposon protein, unclassified	154	0.63	1	38.63%	69.47%
LOC_Os03g40720	genomic|UDP-glucose 6-dehydrogenase	149	0.84	1	45.92%	66.96%
LOC_Os09g23330	genomic|retrotransposon, centromere-specific	144	0.84	1	35.62%	69.05%
LOC_Os01g23160	genomic|retrotransposon protein, unclassified	148	0.84	1	53.22%	62.50%
LOC_Os05g03770	genomic|expressed protein	148	0.85	1	52.36%	62.60%
LOC_Os12g33530	genomic|retrotransposon protein, Ty3-gypsy subclass	148	0.86	1	52.36%	62.40%
LOC_Os06g31000	genomic|retrotransposon protein, unclassified	148	0.87	1	52.36%	62.40%
LOC_Os03g18790	genomic|SAG20, putative, expressed	145	0.93	1	81.12%	58.25%

Table S4.2 Primers used in this study.

Primers	**Sequence**	**Note**
pUbi	TGATATACTTGGATGATGGCA	*Ubi*$_{pro}$*-AT9* genotyping
Pri452-AT9	CCGCAGGTGAATTTGGTGAT	
PDKF1	CTTCGTCTTACACATCACTTGT	*RNAi-AT9* genotyping
T35S-R1	TCAACACATGAGCGAAACCC	
UBIL-F1	GCCCTGCCTTCATACGCTATT	*RNAi-AT9* genotyping
cat-R1	GCTTGCGCTGCAGTTATCAT	
Hyg-F1	TCCACTATCGGCGAGTAC	*Ubi*$_{pro}$*-AT9* and *RNAi-AT9* genotyping
Hyg-R1	CACTGGCAAACTGTGATGGACGAC	
Os01g09010_832F	CACCTGCTGAAGCTGGACAG	LOC_Os01g09010 qPCR
Os01g09010_929R	TCCATCACCGACGACGACAGCA	
CC55 R1-5'	AAGGAGAAAGCCGAACAACG	LOC_Os04g35910 qPCR
CC55 RT1-3'	TCCTCAAGTTTCTTCCTGTAGGC	
UBQ5 RT 1-5'	ACCACTTCGACCGCCACTACT	LOC_Os01g22490 qPCR
UBQ5 RT1-3'	ACGCCTAAGCCTGCTGGTT	
AT9-attBF1	GGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGGGGACGGGGAGCTTCAAG	LOC_Os01g09010 cloning
AT9-attBR1	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGGCAGGTCCTCCTTCATGCCGCG	
GWR-AT9-F	CACCAGTGCAAATCACGTTCATTGTTTC	LOC_Os01g09010 5'-UTR cloning
GWR-AT9-R	TGGCCTTAAATGGAGCGG	

Table S5.1 Sequencing and Mapping Quality.

Name	Reads	Q20 (%)	Mapping rate	Gene (%) reads>0	Gene (%) reads>10
421E4_W1	47339332	97.47%	89.2%	66.7%	46.0%
421E4_W2	46978552	97.72%	89.4%	67.3%	46.5%
421E4_W3	40186449	97.78%	89.3%	66.4%	45.4%
421_L1	43196590	97.52%	89.6%	63.3%	42.2%
421_L2	39916490	97.44%	87.2%	61.8%	40.6%
421_L3	39573803	97.56%	89.3%	64.0%	43.2%
421_SH1	37264955	97.85%	89.0%	64.1%	43.3%
421_SH2	35041320	97.76%	89.1%	65.0%	43.7%
421_SH3	36548274	97.81%	89.3%	64.5%	43.7%
421_SS1	37124262	97.85%	88.1%	64.2%	44.3%
421_SS2	35722847	97.87%	89.2%	64.7%	44.5%
421_SS3	36658424	97.56%	89.2%	61.9%	41.0%
514E4_W1	48506070	97.81%	89.1%	66.7%	46.3%
514E4_W2	47099054	97.83%	89.1%	66.4%	45.7%
514E4_W3	41026324	97.92%	88.9%	65.7%	44.8%
514_L1	49034093	98.03%	89.0%	61.2%	40.1%
514_L2	45326162	98.04%	89.4%	61.2%	40.0%
514_L3	38559126	98.10%	89.2%	60.2%	39.1%
514_SH1	44596974	97.49%	89.2%	62.9%	42.7%
514_SH2	45416549	97.44%	88.8%	65.6%	45.9%
514_SH3	44137172	97.41%	84.2%	63.0%	42.1%
514_SS1	44062260	97.73%	89.1%	63.7%	43.5%
514_SS2	44302692	97.62%	89.1%	62.5%	42.2%
514_SS3	41365192	98.05%	89.1%	63.3%	43.1%
530E4_W1	45683292	97.43%	89.1%	66.5%	45.4%
530E4_W2	44586907	97.28%	89.5%	66.1%	45.2%
530E4_W3	46814212	97.35%	89.5%	66.5%	45.7%
530_L1	36451907	97.94%	89.1%	64.0%	43.1%
530_L2	36356734	97.86%	89.3%	62.7%	40.9%
530_L3	35203037	97.91%	89.2%	63.1%	41.7%
530_SH1	49797217	97.72%	89.1%	67.5%	47.2%
530_SH2	47071555	97.60%	89.1%	65.8%	45.1%
530_SH3	36901860	97.88%	89.2%	63.1%	42.1%
530_SS1	37828915	97.91%	89.0%	65.2%	45.2%
530_SS2	34836285	97.77%	89.3%	64.7%	44.5%
530_SS3	39540941	97.86%	89.4%	63.7%	42.5%
541E4_W1	44371169	97.72%	88.9%	66.5%	45.2%
541E4_W2	46244076	97.77%	89.5%	66.7%	45.7%
541E4_W3	44875759	97.83%	89.0%	66.5%	45.2%
541_L1	47290966	97.30%	89.5%	60.7%	39.4%

541_L2	50325891	97.27%	88.9%	60.6%	39.3%
541_L3	49528661	97.53%	89.8%	61.9%	40.4%
541_SH1	44597546	97.66%	89.1%	63.3%	42.6%
541_SH2	47273129	97.70%	89.1%	66.6%	45.9%
541_SH3	47454354	97.68%	89.1%	64.7%	43.9%
541_SS1	45284383	97.78%	89.1%	63.8%	43.5%
541_SS2	46252597	97.43%	88.9%	64.8%	44.6%
541_SS3	45941383	97.40%	89.1%	63.2%	42.2%
Average	**42906161**	**97.69%**	**89.0%**	**64.3%**	**43.5%**
Max	**50325891**	**98.10%**	**89.8%**	**67.5%**	**47.2%**
Min	**34836285**	**97.27%**	**84.2%**	**60.2%**	**39.1%**

Table S5.2. Top 50 up- and down-regulated DEGs in WRvsWD.

ID	log2(Fold change)	Annotation
Up-regulated		
Pavir.8KG265300	19.42	OsSBeL1 - Putative Serine Beta-Lactamase homologue, expressed
Pavir.8KG138500	19.27	ATPase BadF/BadG/BcrA/BcrD type
Pavir.8KG118800	11.93	expressed protein
Pavir.3KG406600	10.92	
Pavir.2NG158600	10.16	
Pavir.6KG223700	9.62	NB-ARC domain containing protein, expressed
Pavir.2NG413600	9.45	pentatricopeptide
Pavir.9NG050600	9.18	
Pavir.4NG023700	9.06	expressed protein
Pavir.J790200	8.99	stripe rust resistance protein Yr10
Pavir.8KG199500	8.98	transposon protein, putative, CACTA, En/Spm sub-class, expressed
Pavir.1NG517100	8.92	
Pavir.7NG082600	8.84	seven in absentia protein family protein, expressed
Pavir.3KG451300	8.81	expressed protein
Pavir.2KG257500	8.79	
Pavir.6KG223800	8.54	NB-ARC domain containing protein, expressed
Pavir.5KG185100	8.42	
Pavir.J055100	8.29	tropinone reductase 2
Pavir.9KG169300	8.26	
Pavir.7NG423400	8.19	
Pavir.2NG176400	8.17	ubiquitin family protein
Pavir.3NG055200	8.14	
Pavir.J684000	8.07	
Pavir.8KG315000	8.02	
Pavir.7KG217800	7.94	retrotransposon protein, putative, unclassified, expressed
Pavir.5NG009800	7.93	
Pavir.5NG483900	7.90	
Pavir.J045700	7.87	
Pavir.9KG640500	7.86	
Pavir.8KG300300	7.81	ADP-ribosylation factor
Pavir.1NG534900	7.71	transposon protein, putative, Pong sub-class, expressed
Pavir.J398700	7.71	
Pavir.2NG255400	7.69	

Pavir.J180500	7.68	OsFBX338 - F-box domain containing protein, expressed
Pavir.8KG255800	7.65	transposon protein, putative, Pong sub-class, expressed
Pavir.2NG257500	7.60	PIF-like orf1
Pavir.1KG302600	7.57	ABC transporter, ATP-binding protein
Pavir.3KG467000	7.57	
Pavir.2NG413700	7.53	pentatricopeptide
Pavir.J090800	7.48	
Pavir.J345600	7.42	
Pavir.1KG302800	7.35	Cysteine-rich receptor-like protein kinase 21 precursor
Pavir.4NG336500	7.34	expressed protein
Pavir.5KG360900	7.32	
Pavir.2KG143700	7.27	receptor-like protein kinase 5 precursor
Pavir.7NG115100	7.23	
Pavir.8NG269500	7.21	
Pavir.6NG177800	7.18	
Pavir.9NG415100	7.18	
Pavir.8KG071500	7.10	carrier
Down-regulated		
Pavir.1KG258900	-29.31	
Pavir.2NG208000	-27.06	F-box/LRR-repeat protein 14
Pavir.2NG198100	-18.71	
Pavir.J143000	-16.97	expressed protein
Pavir.8KG306800	-10.28	
Pavir.5NG568100	-10.09	
Pavir.1NG087600	-9.75	
Pavir.2NG134600	-9.32	RGH1A
Pavir.3KG287500	-9.27	expressed protein
Pavir.8KG338800	-9.15	resistance protein
Pavir.2KG171700	-9.00	
Pavir.J584600	-8.70	
Pavir.5NG628500	-8.54	
Pavir.2NG224500	-8.53	
Pavir.J771500	-8.34	
Pavir.2KG200600	-8.34	transposon protein, putative, Pong sub-class, expressed
Pavir.3KG293000	-8.30	expressed protein
Pavir.8KG369100	-8.18	
Pavir.5NG524200	-8.16	
Pavir.2NG225800	-8.15	
Pavir.3NG238000	-8.13	expressed protein

Pavir.3KG287600	-8.01	
Pavir.J462800	-7.98	NB-ARC domain containing protein, expressed
Pavir.2KG150600	-7.96	retrotransposon protein, putative, unclassified, expressed
Pavir.J551800	-7.93	
Pavir.8KG294800	-7.85	plant protein of unknown function domain containing protein, expressed
Pavir.5NG625000	-7.84	glycosyl hydrolases family 17
Pavir.J172200	-7.78	
Pavir.2KG271500	-7.77	Plant PDR ABC transporter associated domain containing protein, expressed
Pavir.3KG237000	-7.65	
Pavir.J255100	-7.56	
Pavir.1KG131900	-7.56	calmodulin binding protein
Pavir.2NG163900	-7.47	
Pavir.3KG432500	-7.43	
Pavir.J059300	-7.43	expressed protein
Pavir.2KG593900	-7.35	
Pavir.J769900	-7.29	
Pavir.3KG389300	-7.25	REX1 DNA Repair family protein, expressed
Pavir.2KG271700	-7.24	Plant PDR ABC transporter associated domain containing protein, expressed
Pavir.2NG358800	-7.24	RNA recognition motif containing protein
Pavir.J013400	-7.23	cytochrome P450
Pavir.3KG153600	-7.12	expressed protein
Pavir.5NG533800	-7.12	pentatricopeptide containing protein
Pavir.1NG088600	-7.11	expressed protein
Pavir.8KG354900	-7.09	stripe rust resistance protein Yr10
Pavir.7KG051500	-7.09	
Pavir.2KG130200	-7.05	
Pavir.2KG269700	-6.99	triacylglycerol lipase 1 precursor
Pavir.2NG068200	-6.86	
Pavir.6KG139400	-6.85	

Table S5.3. Top 20 dominant genes in stem and leaf (normalized counts)

	ID	Average counts	annotation
Top 20 in stem			
	Pavir.9NG522800	66454	sucrose synthase
	Pavir.1KG386100	62663	phenylalanine ammonia-lyase
	Pavir.3KG526100	51773	tubulin/FtsZ domain containing protein
	Pavir.9NG179000	44866	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
	Pavir.9KG389000	40968	sucrose synthase
	Pavir.9NG124500	40156	tubulin/FtsZ domain containing protein
	Pavir.2KG395800	36391	chlorophyll A-B binding protein
	Pavir.3KG131900	36019	S-adenosylmethionine synthetase
	Pavir.J553400	30157	chlorophyll A-B binding protein
	Pavir.2NG462200	29685	expressed protein
	Pavir.2NG015800	28931	S-adenosylmethionine synthetase
	Pavir.3KG555300	26455	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
	Pavir.3KG309300	25377	chlorophyll A-B binding protein
	Pavir.1NG430900	25310	erythronate-4-phosphate dehydrogenase
	Pavir.6NG060500	24939	O-methyltransferase
	Pavir.1NG229200	24893	ribulose bisphosphate carboxylase small chain, chloroplast precursor
	Pavir.5NG344500	23578	S-adenosylmethionine synthetase
	Pavir.7NG355500	22906	phenylalanine ammonia-lyase
	Pavir.9NG220200	21716	chlorophyll A-B binding protein
	Pavir.7KG238100	20918	
Top 20 in leaf			
	Pavir.4KG238300	271305	pyruvate, phosphate dikinase, chloroplast precursor
	Pavir.4KG247000	235888	phosphoenolpyruvate carboxylase
	Pavir.3KG261700	212968	pyruvate, phosphate dikinase, chloroplast precursor
	Pavir.5NG403500	121572	bile acid sodium symporter family protein
	Pavir.5NG263800	111457	carbonic anhydrase, chloroplast precursor
	Pavir.5KG453100	102653	carbonic anhydrase, chloroplast precursor
	Pavir.8KG393900	88262	metallothionein
	Pavir.2KG395800	81681	chlorophyll A-B binding protein
	Pavir.6KG096700	70851	photosystem II 10 kDa polypeptide, chloroplast precursor
	Pavir.2NG462200	68328	expressed protein
	Pavir.J651100	59749	chlorophyll A-B binding protein
	Pavir.1KG064100	59063	

	Pavir.5KG445100	57502	bile acid sodium symporter family protein
	Pavir.9NG220200	57166	chlorophyll A-B binding protein
	Pavir.9KG550900	56332	glyceraldehyde-3-phosphate dehydrogenase
	Pavir.8NG306600	54022	metallothionein
	Pavir.J165300	52210	glyceraldehyde-3-phosphate dehydrogenase
	Pavir.5KG001400	51887	DUF250 domain containing protein
	Pavir.5NG089500	50345	metallothionein
	Pavir.4NG299500	50267	OsFtsH2 FtsH protease, homologue of AtFtsH2/8, expressed

Table S5.4. Top 20 leaf and stem specific genes.

ID	Normalized Counts	Annotation
Leaf-specific		
Pavir.4KG238300	271306	pyruvate, phosphate dikinase, chloroplast precursor
Pavir.4KG247000	235889	phosphoenolpyruvate carboxylase
Pavir.3KG261700	212969	pyruvate, phosphate dikinase, chloroplast precursor
Pavir.5NG403500	121573	bile acid sodium symporter family protein
Pavir.5NG263800	111458	carbonic anhydrase, chloroplast precursor
Pavir.5KG453100	102654	carbonic anhydrase, chloroplast precursor
Pavir.8KG393900	88262	metallothionein
Pavir.2KG395800	81681	chlorophyll A-B binding protein
Pavir.6KG096700	70852	photosystem II 10 kDa polypeptide, chloroplast precursor
Pavir.2NG462200	68329	expressed protein
Pavir.J651100	59749	chlorophyll A-B binding protein
Pavir.1KG064100	59063	none
Pavir.5KG445100	57503	bile acid sodium symporter family protein
Pavir.9NG220200	57166	chlorophyll A-B binding protein
Pavir.9KG550900	56332	glyceraldehyde-3-phosphate dehydrogenase
Pavir.8NG306600	54022	metallothionein
Pavir.J165300	52210	glyceraldehyde-3-phosphate dehydrogenase
Pavir.5KG001400	51888	DUF250 domain containing protein
Pavir.5NG089500	50346	metallothionein
Pavir.4NG299500	50267	OsFtsH2 FtsH protease, homologue of AtFtsH2/8
Stem-specific		
Pavir.9NG522800	66454	sucrose synthase
Pavir.1KG386100	62663	phenylalanine ammonia-lyase
Pavir.3KG526100	51773	tubulin/FtsZ domain containing protein
Pavir.9NG179000	44866	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
Pavir.9KG389000	40968	sucrose synthase
Pavir.9NG124500	40156	tubulin/FtsZ domain containing protein
Pavir.2KG395800	36391	chlorophyll A-B binding protein
Pavir.3KG131900	36019	S-adenosylmethionine synthetase
Pavir.J553400	30157	chlorophyll A-B binding protein
Pavir.2NG462200	29685	none
Pavir.2NG015800	28931	S-adenosylmethionine synthetase
Pavir.3KG555300	26455	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
Pavir.3KG309300	25377	chlorophyll A-B binding protein
Pavir.1NG430900	25310	erythronate-4-phosphate dehydrogenase
Pavir.6NG060500	24939	O-methyltransferase

Pavir.1NG229200	24893	ribulose bisphosphate carboxylase small chain, chloroplast precursor
Pavir.5NG344500	23578	S-adenosylmethionine synthetase
Pavir.7NG355500	22906	phenylalanine ammonia-lyase
Pavir.9NG220200	21716	chlorophyll A-B binding protein
Pavir.7KG238100	20918	none

Table S5.5. Top 50 of up and down-regulated DEGs of SH/SS comparison.

ID	log2 (Fold change)	Annotation
Up-regulated		
Pavir.2NG180500	5.4	hydroquinone glucosyltransferase: GT1
Pavir.6KG130600	5.3	cytokinin-O-glucosyltransferase 2: GT1
Pavir.4KG081800	5.1	UDP-glucoronosyl and UDP-glucosyl transferase: GT1
Pavir.1NG203000	5.0	UDP-glucoronosyl/UDP-glucosyl transferase: GT1
Pavir.7KG227900	5.9	cytochrome P450: F5H
Pavir.2KG270900	5.2	cytochrome P450: F5H
Pavir.1KG215200	12.8	leucine-rich
Pavir.2KG384300	7.1	ribonuclease T2 family domain containing protein, expressed
Pavir.1KG145100	6.9	receptor-like protein kinase 5 precursor
Pavir.J314700	6.3	app1
Pavir.5KG284600	6.2	osFTL12 FT-Like12 homologous to Flowering Locus T gene; contains Pfam profile PF01161: Phosphatidylethanolamine-binding protein, expressed
Pavir.2KG384500	6.2	none
Pavir.8KG172300	6.2	dirigent
Pavir.1KG287400	6.0	receptor-like protein kinase 5 precursor
Pavir.2KG510200	6.0	transporter family protein
Pavir.8KG039600	5.9	leucoanthocyanidin dioxygenase
Pavir.1NG421700	5.7	uncharacterized membrane protein
Pavir.2KG384400	5.6	ribonuclease T2 family domain containing protein, expressed
Pavir.4NG266100	5.6	
Pavir.3KG432700	5.6	OsWAK46- OsWAK receptor-like protein kinase, expressed
Pavir.6NG135600	5.5	terpene synthase
Pavir.3NG260400	5.5	cysteine-rich receptor-like protein kinase 8 precursor
Pavir.5KG007900	5.4	C4-dicarboxylate transporter/malic acid transport protein, expressed
Pavir.5NG023800	5.4	expressed protein
Pavir.2KG291200	5.4	MATE efflux family protein
Pavir.5NG610400	5.4	nodulin MtN3 family protein
Pavir.7KG002000	5.3	wall-associated receptor kinase 3 precursor
Pavir.6KG306900	5.3	transporter family protein
Pavir.9NG358600	5.3	oxidoreductase, aldo/keto reductase family protein
Pavir.8NG324400	5.3	tropinone reductase 2
Pavir.7KG058100	5.3	C-methyltransferase

Pavir.J780600	5.3	receptor-like protein kinase 5 precursor
Pavir.5KG373100	5.2	fatty acid hydroxylase
Pavir.5KG450100	5.2	dihydroflavonol-4-reductase
Pavir.5KG108100	5.2	hydroxylase
Pavir.5NG601500	5.2	MLO domain containing protein
Pavir.5NG017000	5.2	C4-dicarboxylate transporter/malic acid transport protein, expressed
Pavir.7KG034900	5.2	wall-associated receptor kinase 3 precursor
Pavir.J189200	5.1	oxidoreductase, short chain dehydrogenase/reductase family domain containing family, expressed
Pavir.5KG132100	5.1	OsWAK28 - OsWAK receptor-like protein kinase, expressed
Pavir.9KG487900	5.1	cysteine-rich repeat secretory protein 55 precursor
Pavir.9KG572400	5.1	catalase domain containing protein, expressed
Pavir.6NG148300	5.1	verticillium wilt disease resistance protein Ve2
Pavir.8KG334600	5.0	legume lectins beta domain containing protein, expressed
Pavir.9NG651100	5.0	BTBN5 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with non-phototropic hypocotyl 3 NPH3 and coiled-coil domains, expressed
Pavir.3NG205200	5.0	nodulin MtN3 family protein
Pavir.7KG428200	5.0	lipase
Pavir.9NG104700	5.0	expressed protein
Pavir.4NG314200	5.0	OsHKT2;4 - Na+ transporter, expressed
Pavir.8NG337200	5.0	TKL_IRAK_DUF26-lc.10 - DUF26 kinases have homology to DUF26 containing loci, expressed
Down-regulated		
Pavir.7KG416200	-4.2	OsHKT1;1 - Na+ transporter, expressed
Pavir.9NG807900	-3.9	OsIAA20 - Auxin-responsive Aux/IAA gene family member, expressed
Pavir.J349200	-3.0	gibberellin 2-beta-dioxygenase
Pavir.5NG423400	-2.9	expressed protein
Pavir.1KG208300	-2.6	meiotic coiled-coil protein 7
Pavir.8KG336400	-2.5	expressed protein
Pavir.5NG537500	-2.4	protein disulfide isomerase
Pavir.1KG469300	-2.2	MYB family transcription factor
Pavir.5NG175600	-2.1	PPR repeat containing protein, expressed
Pavir.4NG066900	-2.0	expressed protein
Pavir.7NG049600	-1.9	2Fe-2S iron-sulfur cluster binding domain containing protein, expressed
Pavir.1NG292800	-1.8	chloride channel protein
Pavir.5KG634500	-1.5	zinc finger, C3HC4 type domain containing protein, expressed

Pavir.5KG341600	-1.5	ZmGR2c
Pavir.7NG415400	-1.4	phosphatidylinositol-4-phosphate 5-Kinase
Pavir.5KG671400	-1.3	Sad1 / UNC-like C-terminal domain containing protein
Pavir.1NG490500	-1.3	BTB1 - Bric-a-Brac,Tramtrack, Broad Complex BTB domain, expressed
Pavir.7KG423300	-1.3	D-3-phosphoglycerate dehydrogenase, chloroplast precursor
Pavir.6NG328400	-1.3	peroxidase precursor
Pavir.6NG071300	-1.3	oxysterol-binding protein-related protein 6
Pavir.1NG490600	-1.3	expressed protein
Pavir.2NG617600	-1.3	OsCam1-3 - Calmodulin, expressed
Pavir.4NG267500	-1.2	villin protein
Pavir.2NG574400	-1.2	IBS1
Pavir.6KG375200	-1.1	ras-related protein
Pavir.9NG727300	-1.1	zinc finger, C3HC4 type domain containing protein, expressed
Pavir.5KG123100	-1.1	xylosyltransferase
Pavir.9NG446800	-1.1	OsCam1-2 - Calmodulin, expressed
Pavir.2NG554000	-1.1	tubulin/FtsZ domain containing protein
Pavir.6KG004300	-1.1	OsCML7 - Calmodulin-related calcium sensor protein, expressed
Pavir.9NG434800	-1.1	zinc finger DHHC domain-containing protein
Pavir.5KG546600	-1.1	villin protein
Pavir.6NG240200	-1.1	conserved oligomeric Golgi complex component 3
Pavir.1NG470100	-1.0	expressed protein
Pavir.4NG296100	-1.0	tubulin/FtsZ domain containing protein
Pavir.2KG401300	-1.0	phosphatidylinositol 3- and 4-kinase family protein
Pavir.2NG398200	-1.0	phosphatidylinositol 3- and 4-kinase family protein
Pavir.9KG313700	-1.0	oxysterol-binding protein
Pavir.6KG387500	-1.0	exo70 exocyst complex subunit
Pavir.5NG492000	-1.0	villin protein
Pavir.7NG386700	-1.0	glycosyl transferase 8 domain containing protein
Pavir.5NG329100	-1.0	ras-related protein
Pavir.1NG143700	-1.0	protein phosphatase 2C
Pavir.9KG324600	-1.0	expressed protein
Pavir.6NG038700	-1.0	OsCML7 - Calmodulin-related calcium sensor protein, expressed
Pavir.3KG508500	-1.0	lipase class 3 family protein
Pavir.9NG127100	-1.0	adenylyl cyclase-associated protein
Pavir.5NG096600	-1.0	expressed protein
Pavir.J163300	-1.0	WD domain, G-beta repeat domain containing protein, expressed

Pavir.5NG547100	-0.9	cupin, RmlC-type

Appendix B: Supplementary Figures

Figure S5.2. Heatmap of expression pattern of Mitchell clade of BAHD gene family subclade ii in switchgrass based on normalized counts from DESeq2.

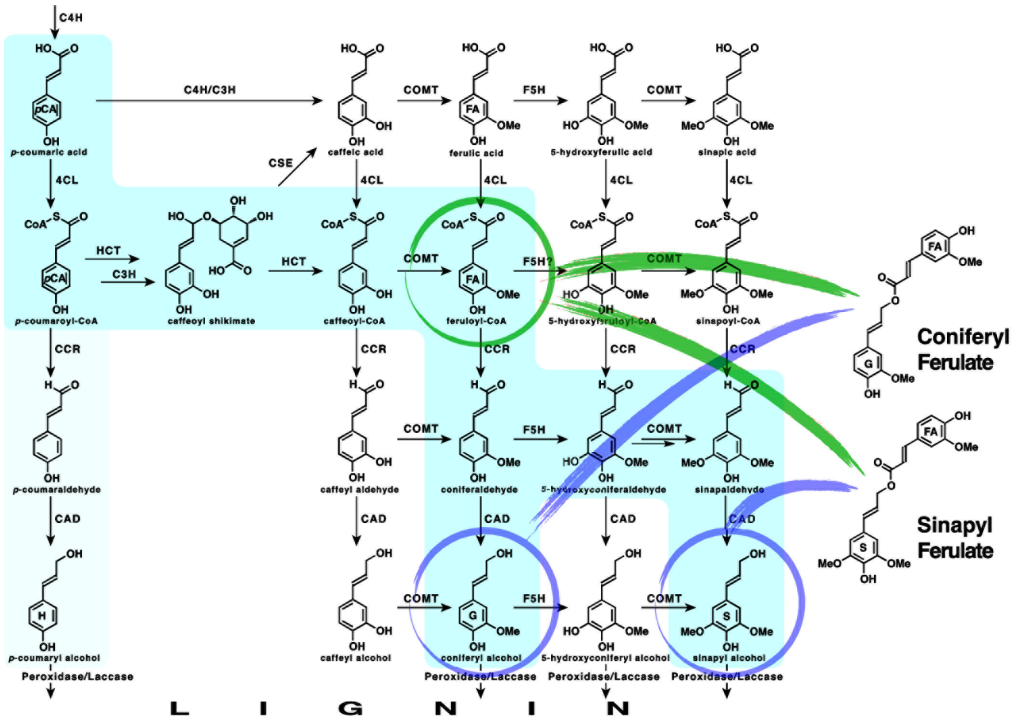
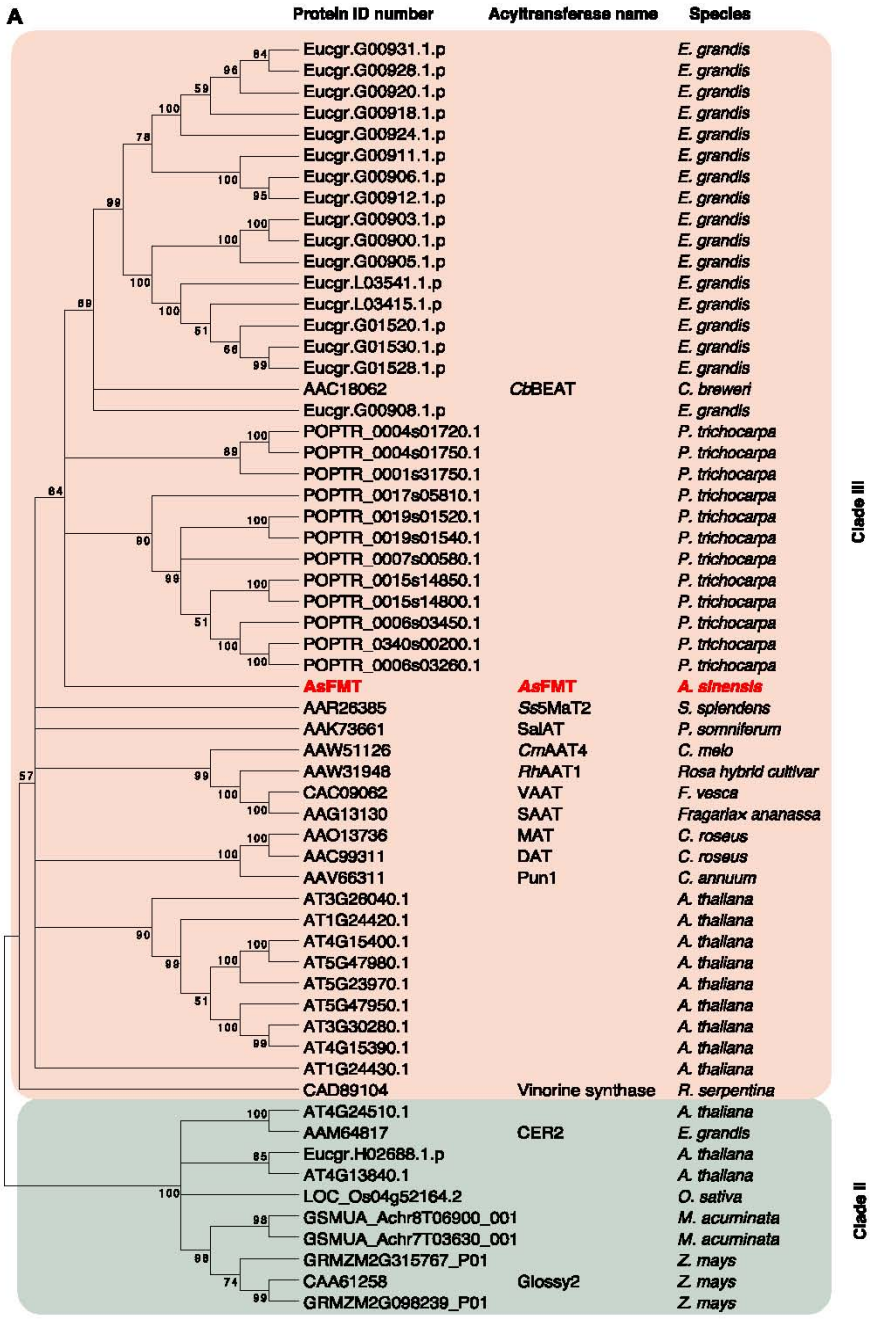
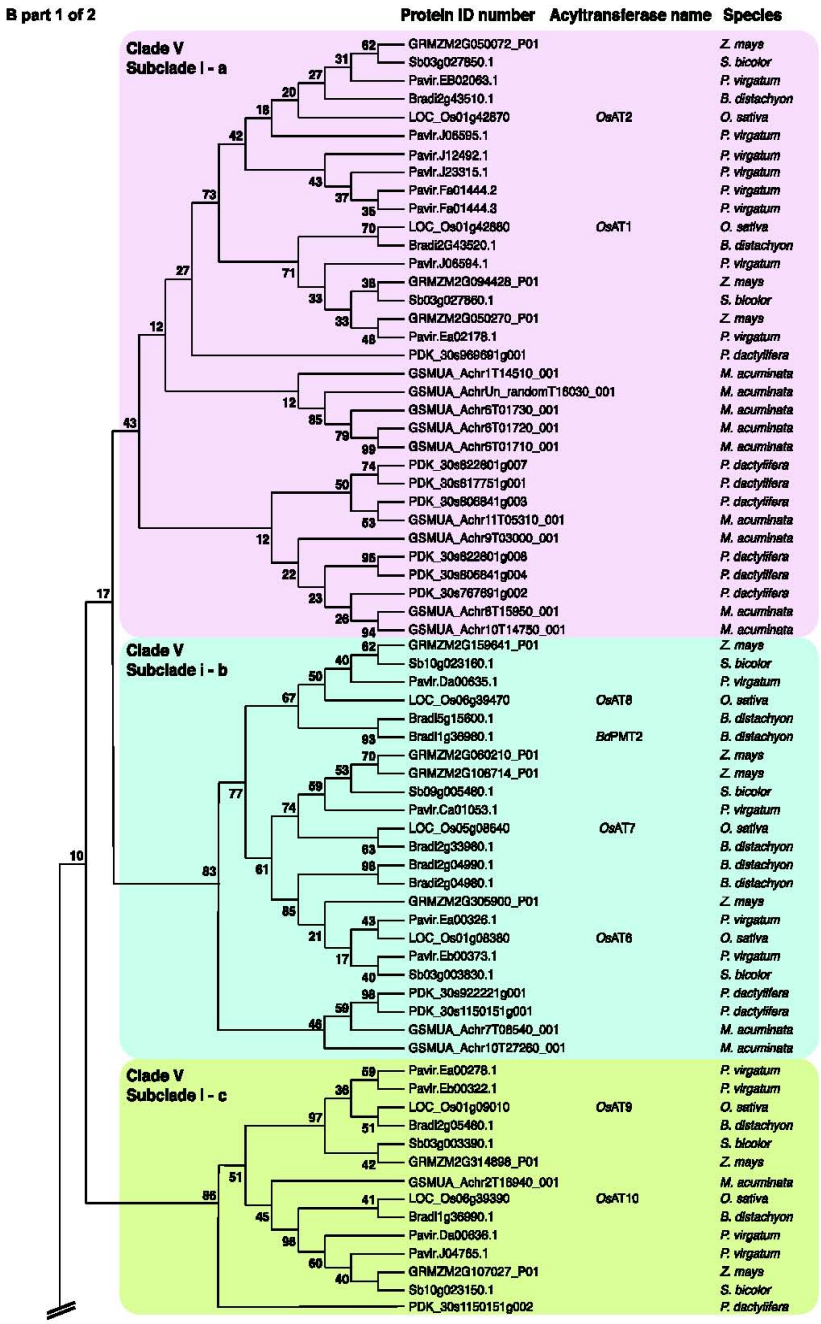


Figure S2.1. The monolignol biosynthetic pathway indicating the formation of ML-FAs. FMT enzymes couple together a monolignol (coniferyl alcohol or sinapyl alcohol) and feruloyl-CoA, which is an intermediate in the pathway, to form ML-FAs. The primary pathway flux in the absence of FMT is indicated using blue shading.



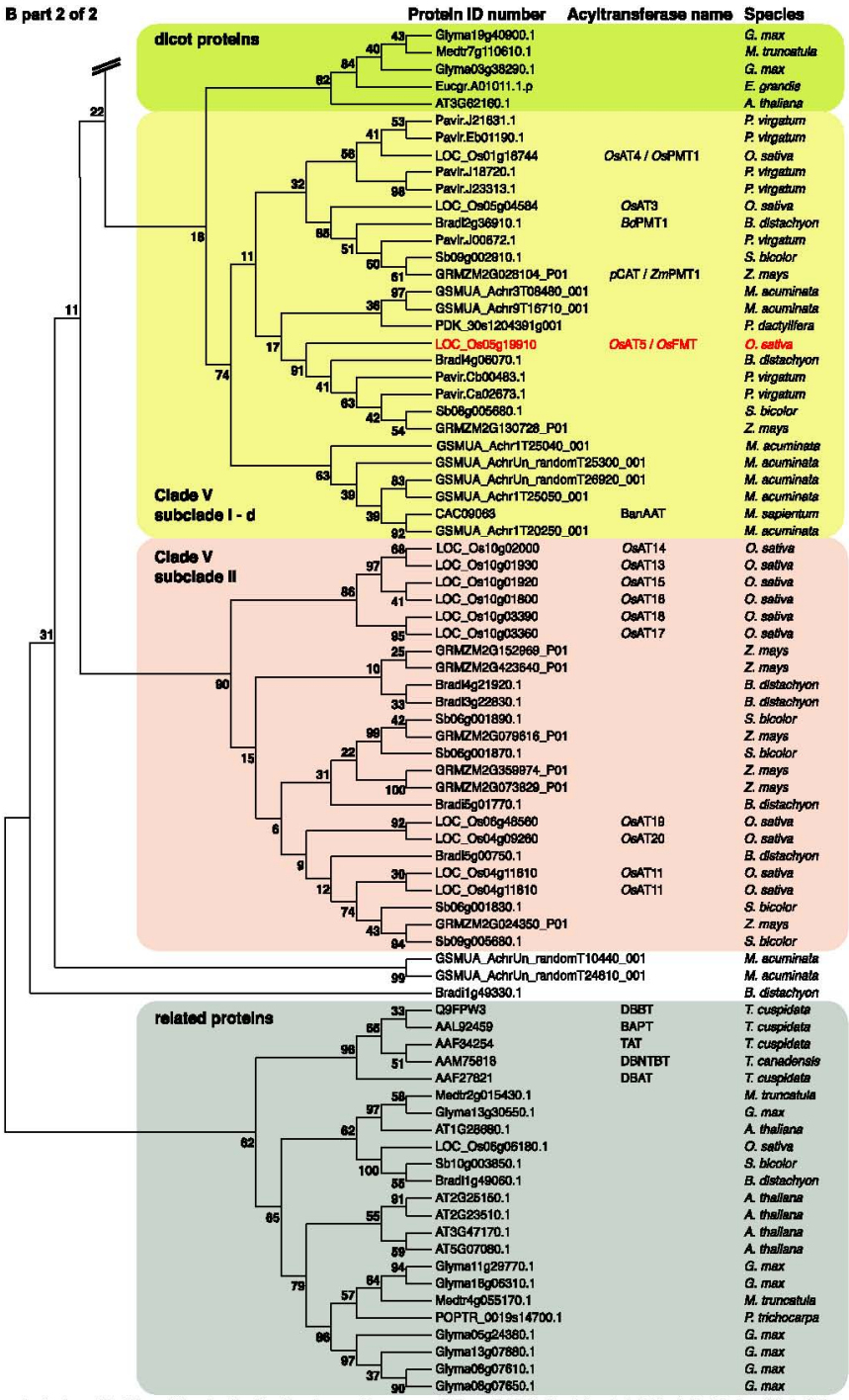
Species codes for locus identifiers with order classifications in parentheses are as follows: AT: *A. thaliana* (Brassicales), Bradi: *B. distachyon* (Poales), Eucgr: *E. grandis* (Myrtales), Glyma: *G. max* (Fabales), GSMUA: *M. acuminata* (Zingiberales), GRMZM: *Z. mays* (Poales), LOC_Os: *O. sativa* (Poales); Medtr: *M. truncatula* (Fabales), Pavir: *P. virgatum* (Poales), PDK: *P. dactylifera* (Arecales); POTR: *P. trichocarpa* (Malpighiales); and Sb: *S. bicolor* (Poales).

Figure S2.2. The phylogenetic reconstruction of BAHD acyl-CoA/ATs is consistent with the convergent evolution of the two feruloyl-CoA/monolignol transferases, OsAT5/FMT and AsFMT. (A) Maximum likelihood phylogeny of AsFMT suggests that grasses do not possess close homologs. Branch values are based on 500 bootstraps.



Species codes for locus identifiers with order classifications in parentheses are as follows: AT: *A. thaliana* (Brassicales), Bradi: *B. distachyon* (Poales), Eucgr: *E. grandis* (Myrtales), Glyma: *G. max* (Fabales), GSMUA: *M. acuminata* (Zingiberales), GRMZM: *Z. mays* (Poales), LOC_Os: *O. sativa* (Poales); Medtr: *M. truncatula* (Fabales), Pavir: *P. virgatum* (Poales), PDK: *P. dactylifera* (Arecales); POTR: *P. trichocarpa* (Malpighiales); and Sb: *S. bicolor* (Poales).

Figure S2.2B 1 of 2. Maximum likelihood phylogeny of Mitchell Clade suggests that grasses and non-grass commelinids have homologs of OsAT5/FMT. Note that this portion of the BAHD phylogenetic tree has no FMT proteins.



Species codes for locus identifiers with order classifications in parentheses are as follows: AT: *A. thaliana* (Brassicales), Bradi: *B. distachyon* (Poales), Eucgr: *E. grandis* (Myrtales), Glyma: *G. max* (Fabales), GSMUA: *M. acuminata* (Zingiberales), GRMZM: *Z. mays* (Poales), LOC_Os: *O. sativa* (Poales); Medtr: *M. truncatula* (Fabales), Pavir: *P. virgatum* (Poales), PDK: *P. dactylifera* (Arecales); POTR: *P. trichocarpa* (Malpighiales); and Sb: *S. bicolor* (Poales).

Figure S2.2B 2 of 2. Maximum likelihood phylogeny of Mitchell Clade suggests that grasses and non-grass commelinids have homologs of OsAT5/FMT.

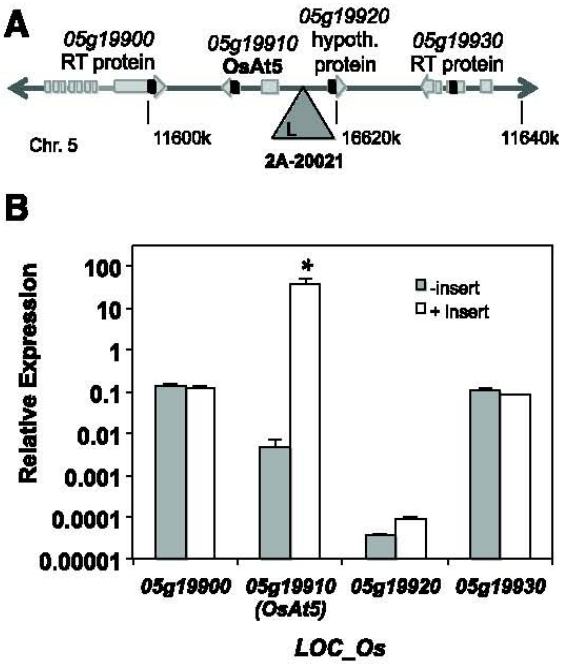


Figure S2.3. Genomic position and gene expression data for the *AT5-D1* rice activation-tagged line. (A) Representation of the rice chromosome near the T-DNA insertion site. Exons are represented by wide bars. The insertion site is represented by the triangle, with the left border, nearest the transcriptional enhancer elements, represented by 'L'. cDNA regions targeted for amplification in qPCR are depicted as black bands. RT stands for retrotransposon and hypoth indicates hypothetical. (B) Average relative gene expression determined via qPCR of RNA isolated from young leaves of homozygous T2 plants with the T-DNA insertion (open) compared with negative segregants (solid) shows that among genes within 20 kbp of the insertion site only At5 expression is altered significantly. The observed minor variations in other nearby genes were not consistent among the 3 biological replicates assayed (not shown). Only 1 of 3 biological replicates for Os05g19920 in the absence of the insert gave a signal distinguishable from background, consistent with that locus' tentative annotation and lack of gene expression evidence. Error bars represent the standard deviation of 3-4 biological replicates. * indicates significantly higher expression ($p<0.01$, Student's t-test).

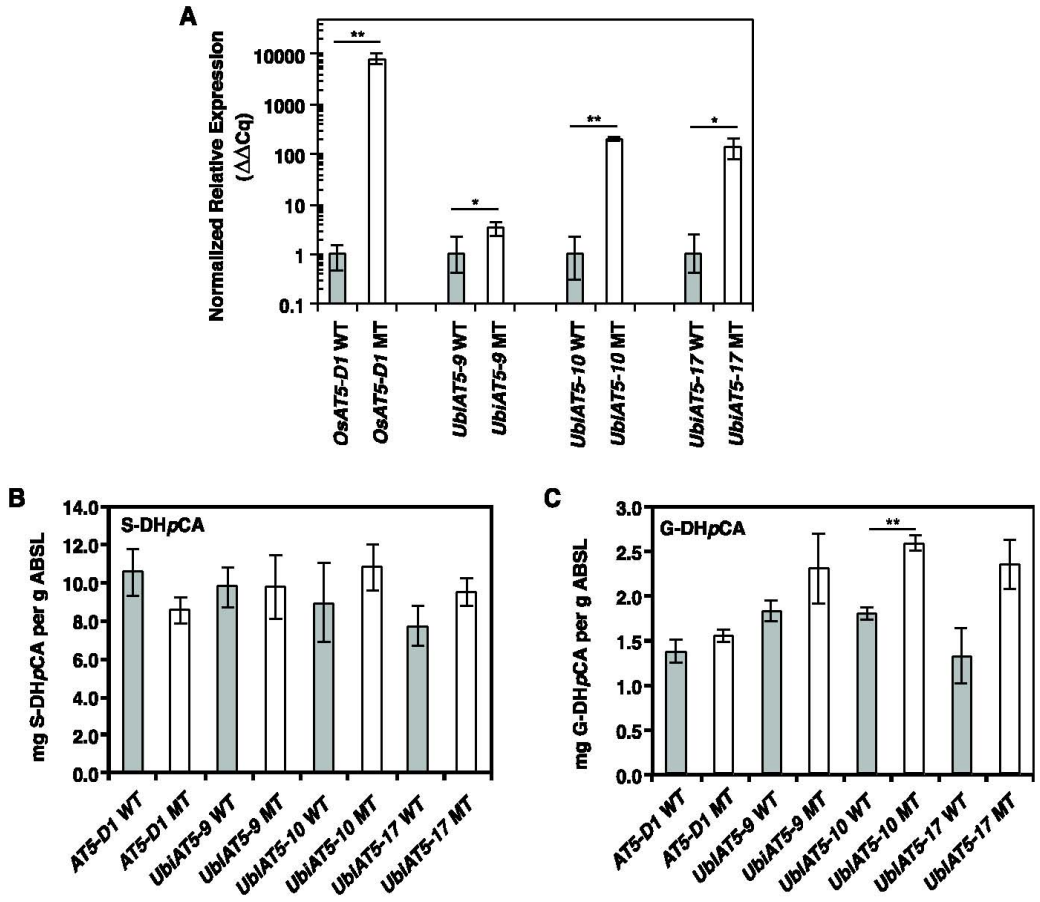


Figure S2.4. *OsAt5* expression data and DFRC-released ML-DH*p*CA conjugates from *OsAT5* rice lines. (**A**) Average normalized gene expression determined via qRT-PCR shows significantly increased *OsAt*5 expression in young leaves of mutant activation tagged *OsAT5-D*1 and *Ubipro::At*5 transgenic lines (open) compared with wild-type, negative segregants (solid). Error bars represent 2*SE of 3-9 biological replicates. (**B-C**) ML-*p*CA conjugates from different over expression lines. Error bars represent the standard error of the mean (SEM) of 3-7 biological replicates, which were measured with technical replicates each. * indicates a difference at $p<0.05$; ** indicates a difference at $0.01<p<0.001$, *** indicates a difference at $p<0.001$ via Student's t test.

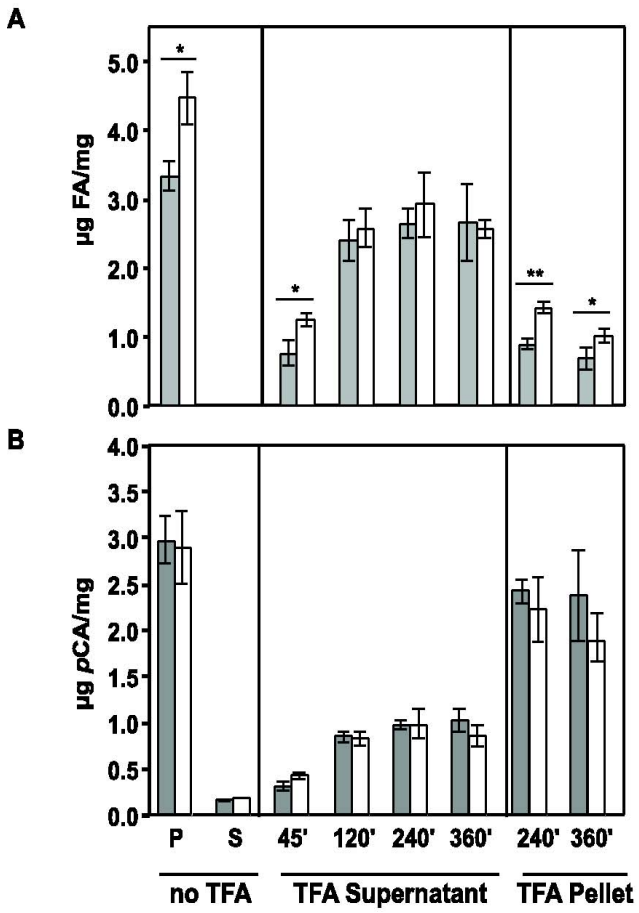


Figure S2.5. The cell wall compositional differences in *OsAT5-D1* straw are predominantly due to the (50 mM TFA, 100°C) insoluble fraction. Wild-type samples are (solid) and mutant samples (open). P indicates the pellet and S the supernatant after TFA treatment (TFA) or mock (no TFA). The numbers indicate the minutes of TFA treatment. (**A**) Ferulic acid (FA) content in dsAIR. (**B**) *p*-Coumaric acid (*p*CA) content in dsAIR. Error bars represent averages of three technical replicates for a pool of biomass from 12 plants of each genotype in the T2 generation. * indicates a difference at $p<0.05$ and ** indicates a difference at $p<0.01$.

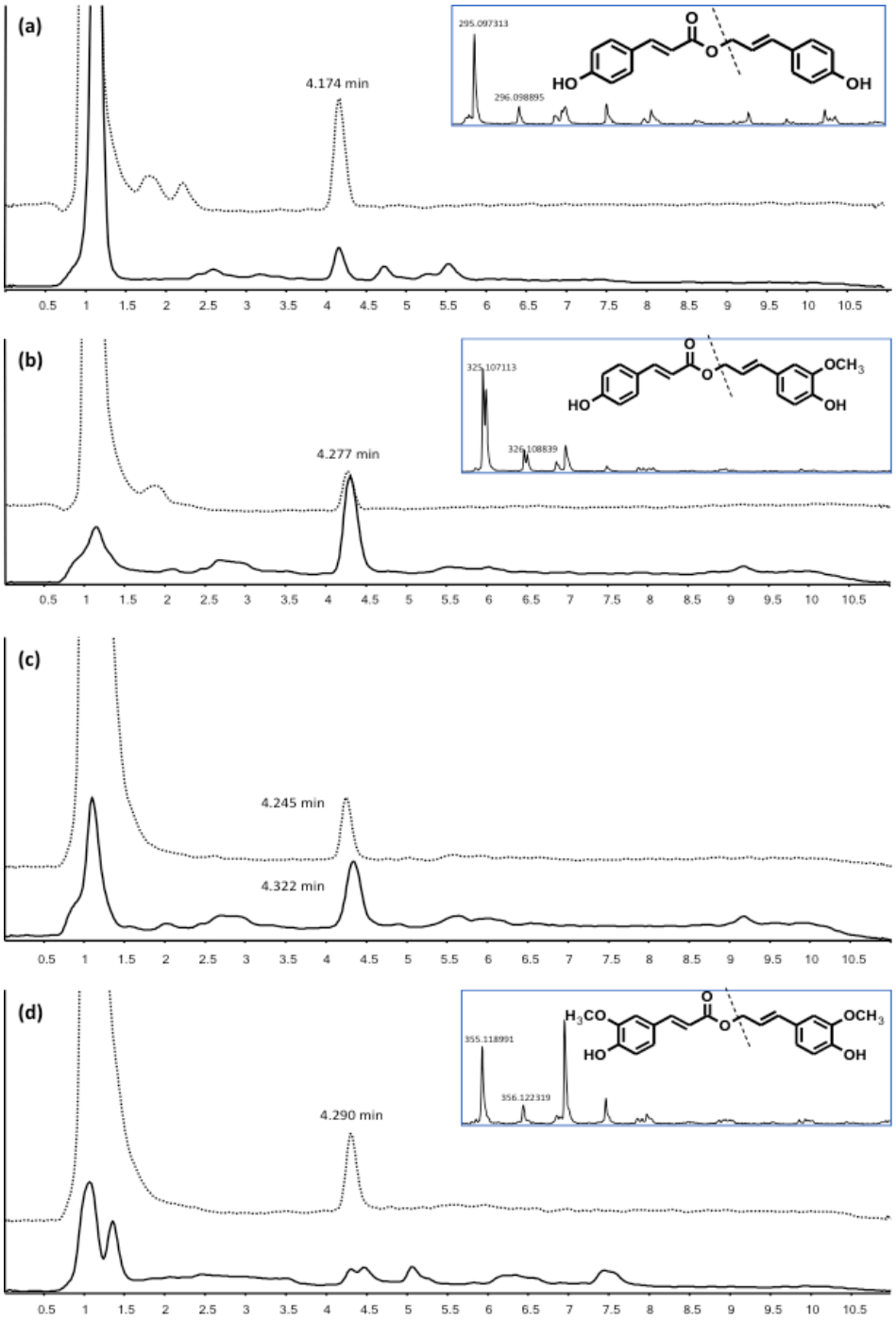


Figure S3.1. LC chromatograms and MS spectra of *pDRf1-4CL5-OsAT4* yeast culture. The four panels indicate yeast cultures fed with different donor/acceptor combinations: (a) coumaric acid and coumaryl alcohol; (b) coumaric acid and coniferyl alcohol; (c) ferulic acid and coumaryl alcohol; and (d) ferulic acid and coniferyl alcohol. The dashed and solid LC chromatogram represents compound with ionization fragment and intact molecules, respectively. The inset represents MS spectrum of the intact molecule.

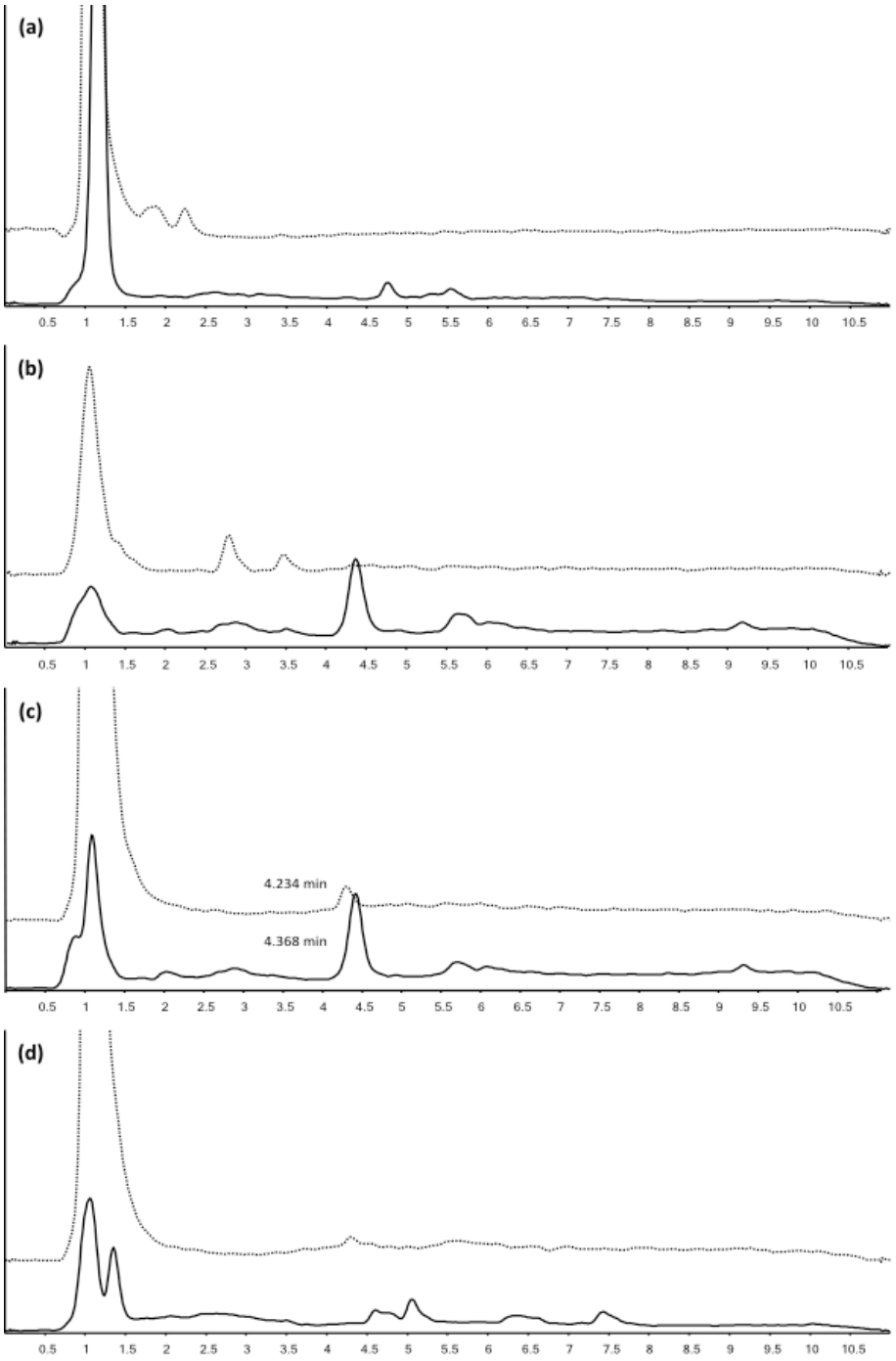


Figure S3.2. LC chromatograms and MS spectra of *pDRf1-4CL5-GW* yeast culture. The four panels indicate yeast cultures fed with different donor/acceptor combinations: (a) coumaric acid and coumaryl alcohol; (b) coumaric acid and coniferyl alcohol; (c) ferulic acid and coumaryl alcohol; and (d) ferulic acid and coniferyl alcohol. The dashed and solid LC chromatogram represents compound with ionization fragment and intact molecule, respectively.

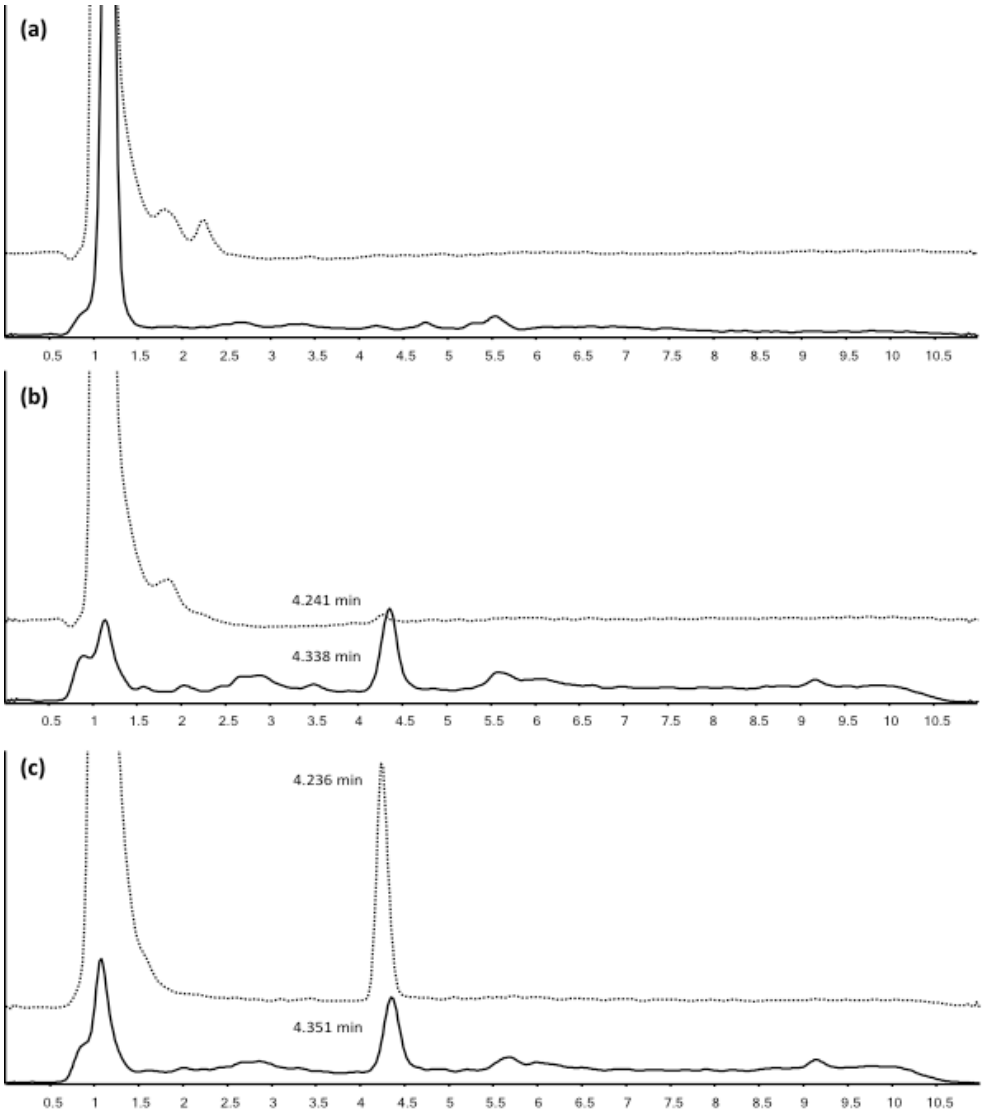


Figure S3.3. LC chromatograms and MS spectra of *pDRf1-4CL5-OsAT5-GW* yeast culture. The four panels indicate yeast cultures fed with different donor/acceptor combinations: (a) coumaric acid and coumaryl alcohol; (b) coumaric acid and coniferyl alcohol; (c) ferulic acid and coumaryl alcohol. The dashed and solid LC chromatogram represents compound with ionization fragment and intact molecule, respectively.

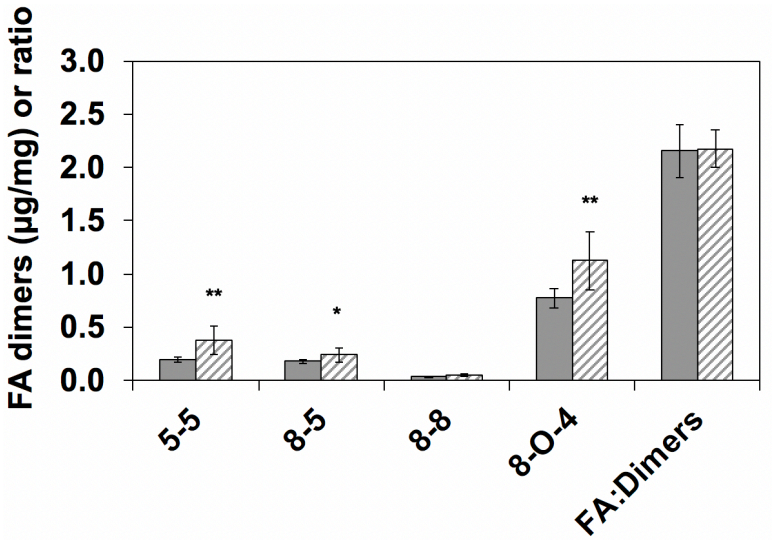


Figure S3.4. The alteration of ester-linked FA-dimers in dsAIR of straws of OsAT5-D1 overexpression lines. The type of FA dimers interlink bonds are 8-O-4 and C-C bonds 5-5, 8-5, 8-8. The data represent mean values (2*SE) from 3 biological replicates. *P < 0.05, **0.001 < P < 0.01, Student's t test.

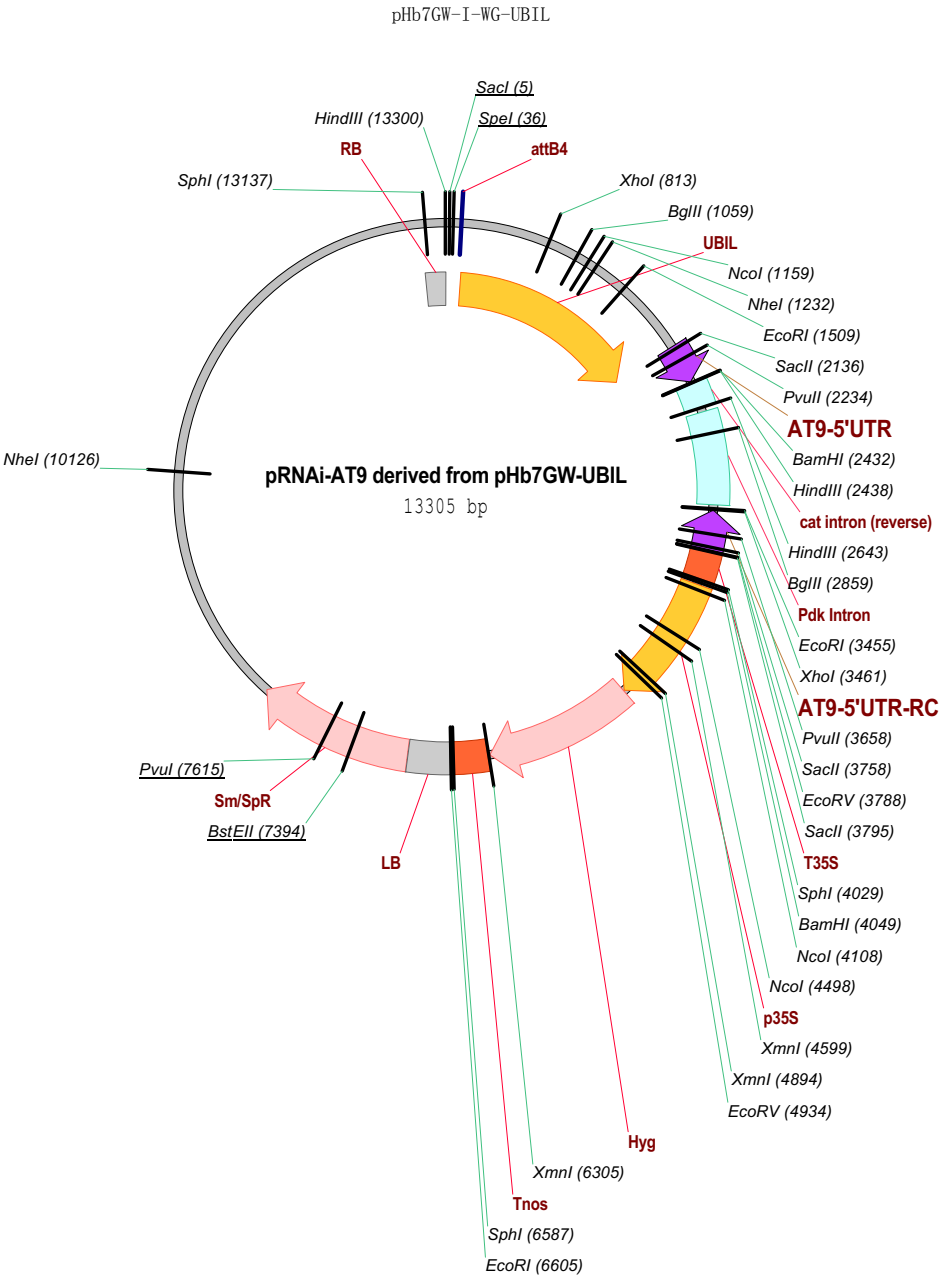
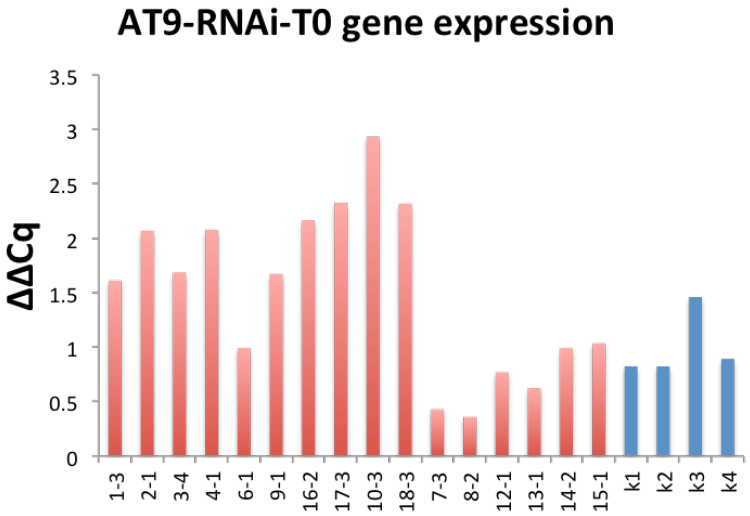
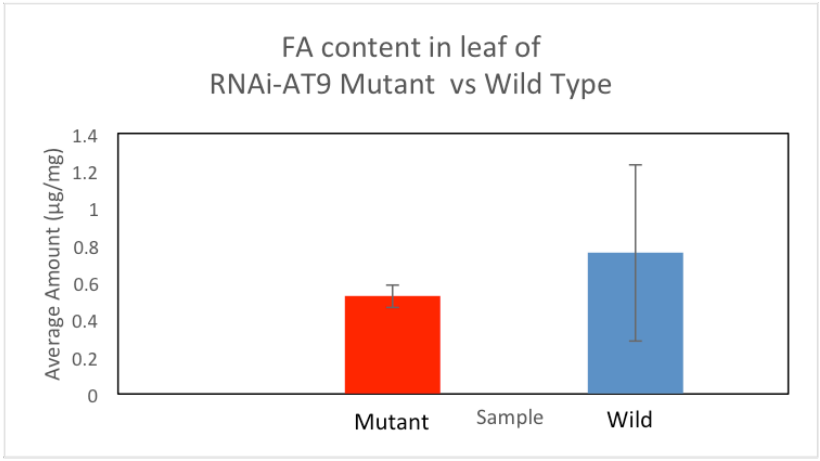


Figure S4.1. RNAi construct design. Purple represents the insertion area with selected RNAi target sequence





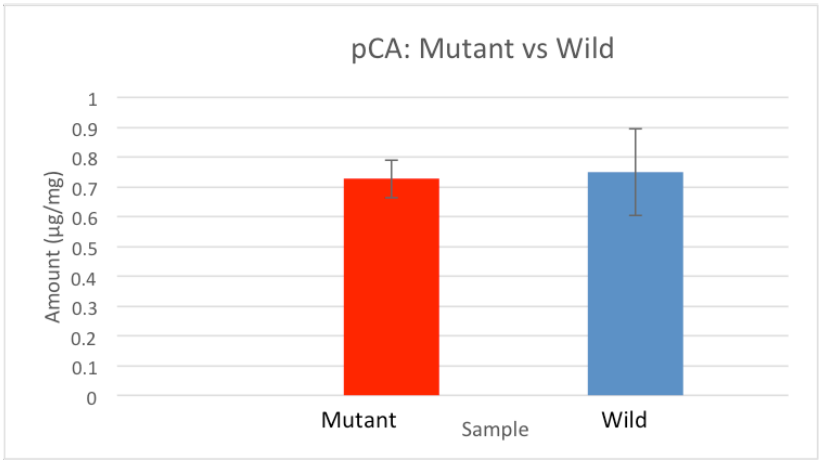


Figure S4.2. Gene expression and wall HCAs from T0 generation of RNAi-AT9 lines.

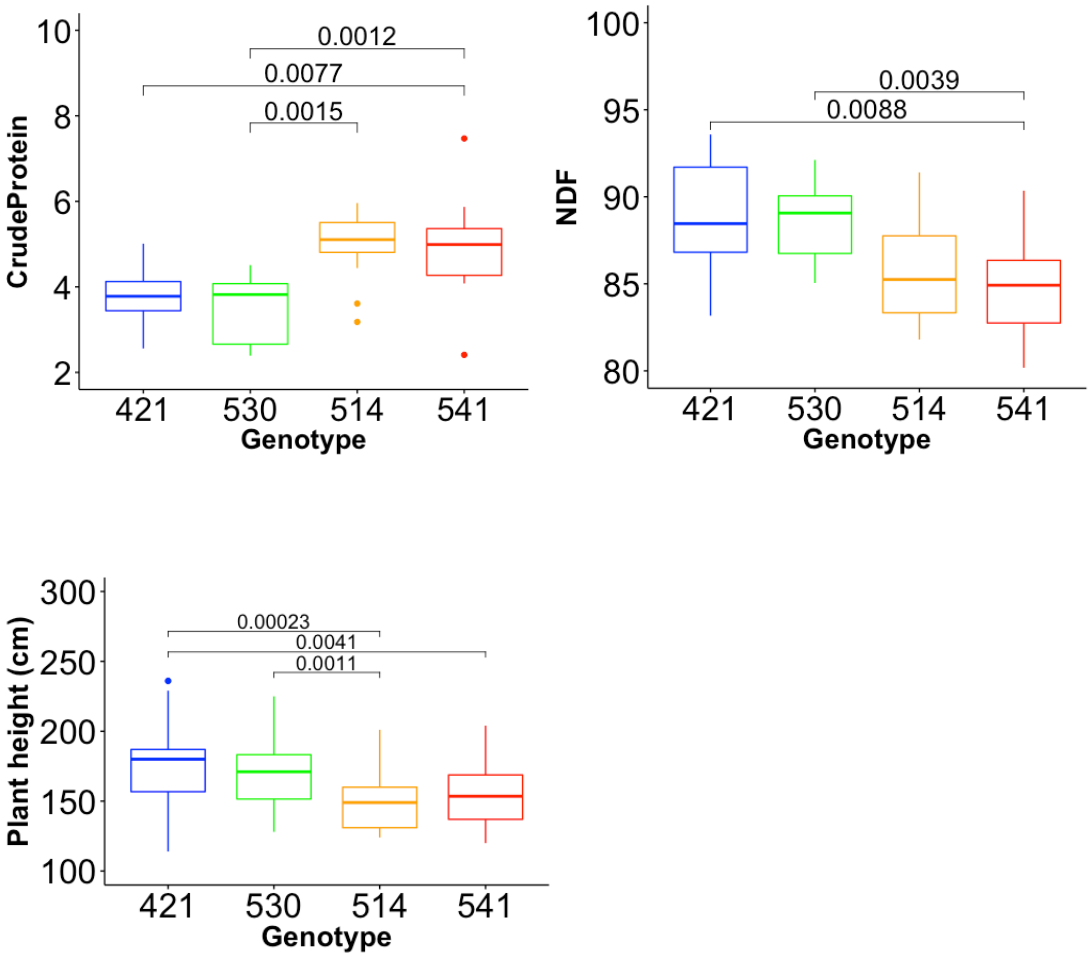


Figure S5.1. Measurements of crude protein, NDF, and plant height from field grown genotypes 421, 530, 514, and 541. Significance p-value was determined by one-way ANOVA.

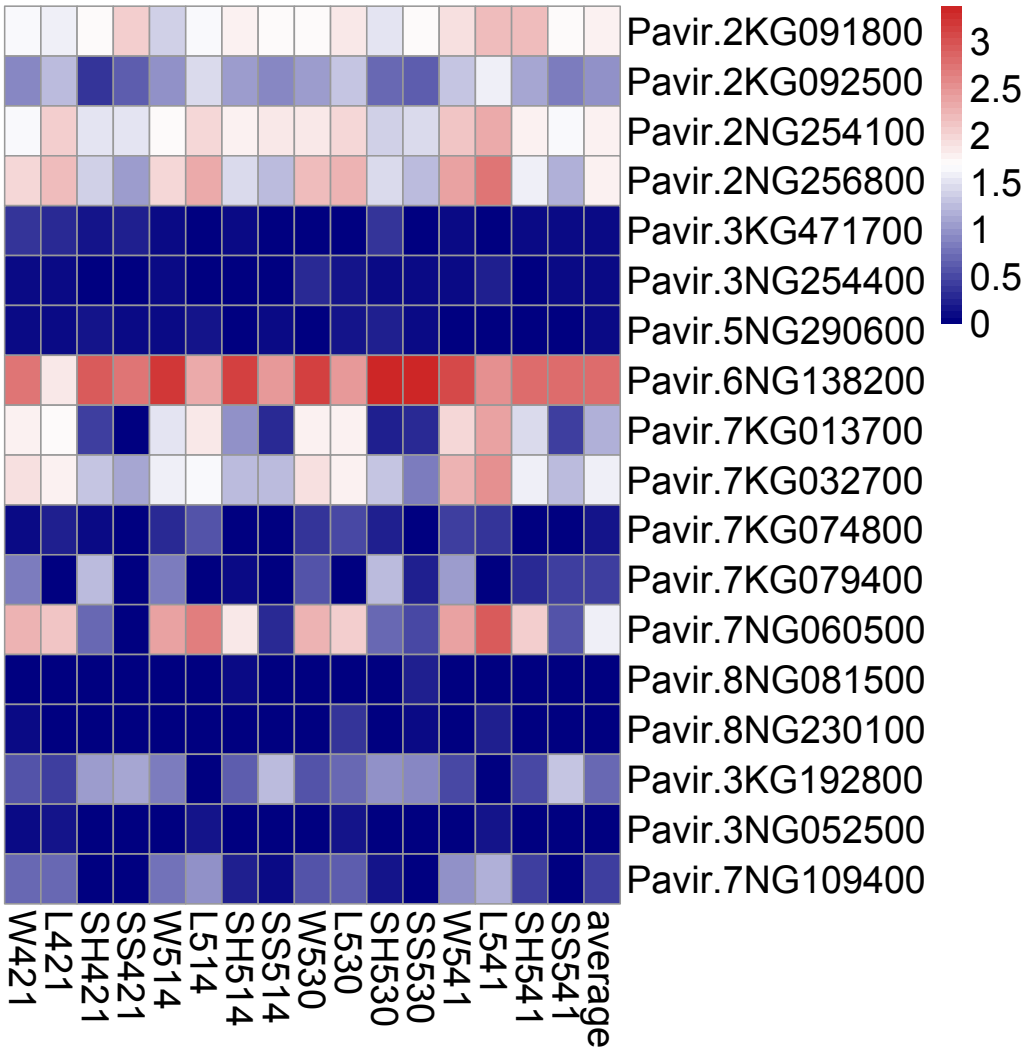


Figure S5.2. **Heatmap of expression pattern of Mitchell clade of BAHD gene family subclade ii in switchgrass based on normalized counts from DESeq2.** From leaf to right are whole tiller (W), leaf (L), hard stem (SH), and soft stem (SS) samples from recalcitrant genotype group (R, 421 and 530) and digestible genotype group (D, 514 and 541).

Appendix C: Supplementary Text

Text S2.1 Sequence of *AsFMT* and *OsAT5*

Text S4.1 Nucleotide sequence of *OsAT9* open reading frame

Text S4.2 Nucleotide sequence of 233-bp *OsAT9* 5'UTR selected for RNAi

Text S2.1 Sequence of AsFMT and OsAT5

Nucleotide sequence of coding region of cDNA clone of *AsFMT*

ATGACGATCATGGAGGTTCAAGTTGTATCTAAGAAGATGGTAAAGCCATCA
GTTCCGACTCCTGACCACCACAAGACTTGCAAATTGACGGCATTCGATCAG
ATTGCTCCTCCGGATCAAGTTCCCATTATTTACTTCTACAACAGCAGCAACA
TCCACAATATTCGCGAGCAATTGGTAAAATCCTTGTCCGAAACTCTAACCAA
GTTTTATCCATTAGCTGGAAGATTTGTTCAAGATGGTTTCTATGTCGATTGTA
ATGATGAAGGGGTCTTGTACGTAGAAGCTGAAGTTAACATTCCGCTAAACG
AATTCATCGGACAAGCAAAGAAAAATATACAACTTATCAATGATCTTGTTC
CGAAAAAAAACTTCAAGGATATTCATTCATATGAAAATCCAATAGTGGGAT
TACAGATGAGTTATTTCAAGTGTGGTGGACTTGCTATTTGCATGTATCTTTC
GCATGTTGTAGCTGATGGATATACAGCAGCAGCATTCACTAAAGAGTGGTC
TAACACAACCAATGGCATCATCAATGGCGATCAACTAGTTTCTTCTTCTCCG
ATTAACTTCGAATTGGCAACTCTAGTCCCAGCTAGAGATTTATCGACGGTGA
TCAAGCCAGCCGTGATGCCACCATCAAAGATCAAGGAAACCAAGGTTGTCA
CAAGGAGGTTTCTGTTCGATGAAAATGCGATATCAGCTTTCAAAGACCATGT
CATCAAATCCGAAAGCGTTAACCGGCCTACACGGGTGGAAGTTGTGACATC
TGTGTTATGGAAGGCTCTGATCAACCAGTCTAAGCTTCCAAGTTCTACACTA
TATTTTCACCTCAACTTTAGAGGGAAAACAGGCATCAACACCCCACCGCTA
GATAATCATTTTTCGCTTTGCGGAAACTTTTACACTCAGGTTCCTACAAGGT
TCAGGGGGGGAAATCAAACAAAACAGGATTTGGAATTGCATGAATTGGTCA
AGTTGTTGAGAGGAAAGTTGCGTAACACTCTGAAGAATTGCTCCGAAATTA
ACACTGCCGATGGGCTGTTCCTGGAAGCAGCTAGTAATTTCAATATTATACA
GGAAGATTTGGAGGACGAACAAGTGGATGTTCGGATTTTTACAACGTTGTG
TAGGATGCCTTTGTATGAAACTGAGTTTGGGTGGGGAAAACCAGAATGGGT
TACCATTCCAGAGATGCATTTGGAGATAGTGTTTCTTTTGGACACTAAATGT
GGGACTGGTATTGAGGCATTAGTGAGCATGGATGAAGCAGATATGCTTCAG
TTTGAACTTGATCCCACCATCTCTGCTTTCGCTTCCTAG

Predicted protein sequence of AsFMT

MTIMEVQVVSKKMVKPSVPTPDHHKTCKLTAFDQIAPPDQVPIIYFYNSSNIHNI
REQLVKSLSETLTKFYPLAGRFVQDGFYVDCNDEGVLYVEAEVNIPLNEFIGQA
KKNIQLINDLVPKKNFKDIHSYENPIVGLQMSYFKCGGLAICMYLSHVVADGYT
AAAFTKEWSNTTNGIINGDQLVSSSPINFELATLVPARDLSTVIKPAVMPPSKIKE
TKVVTRRFLFDENAISAFKDHVIKSESVNRPTRVEVVTSVLWKALINQSKLPSST
LYFHLNFRGKTGINTPPLDNHFSLCGNFYTQVPTRFRGGNQTKQDLELHELVKL
LRGKLRNTLKNCSEINTADGLFLEAASNFNIIQEDLEDEQVDVRIFTTLCRMPLY
ETEFGWGKPEWVTIPEMHLEIVFLLDTKCGTGIEALV
SMDEADMLQFELDPTISAFAS*

Nucleotide sequence of coding region of cDNA clone of *OsFMT* (*OsAT5*)

ATGGTCGCTGTCACCGTGATGAGGAAGTCCCGGAACTTCGTCGGGCCGTCTC
CTCCGACGCCGCCGGCCGAGATCACGACGACGCTAGAGCTGTCGTCCATCG
ACCGCGTGCCCGGGCTGCGCCACAACGTGCGGTCCCTGCACGTGTTCCGCC
GCCACAAGAACAGCGGGCCCGTCGTCGACGGTGATAGCAGGAGGCCGGCC
GCCGTGATCCGCGCGGCGCTCGCCCGGGCGCTGGCGGACTACCCGGCGTTC
GCCGGCCGATTCGTCGGCTCCCTGCTGGCCGGCGACGCCTGCGTCGCGTGCA

CCGGCGAGGGCGCGTGGTTCGTGGAGGCAGCCGCGGACTGCAGCCTCGACG
ACGTGAACGGCCTGGAGTACCCGCTCATGATCTCCGAGGAGGAGCTGCTGC
CTGCCCCCGAGGACGGCGTCGACCCTACCAGTATTCCAGTCATGATGCAGG
TGACTGAATTCACTTGTGGAGGATTTATCTTGGGCCTTGTGGCAGTCCACAC
CCTTGCTGATGGACTTGGAGCAGCACAATTCATCACTGCAGTAGCTGAATTG
GCCCGTGGCATGGACAAGCTCAGGGTGGCTCCCGTGTGGGATCGCTCGCTG
ATACCGAACCCACCTAAGCTCCCTCCTGGGCCACCACCATCGTTCCAGTCCT
TTGGTTTTCAGCATTTCTCCACAGATGTCACCTCTGACCGTATAGCTCACGT
GAAGGCTGAGTACTTCCAGACCTTTGGCCAGTATTGTTCCACCTTTGATGTT
GCTACTGCTAAGGTTTGGCAGGCCAGGACACGGGCCGTCGGGTACAAACCG
GAGATCCAGGTCCATGTGTGTTTCTTTGCAAACACTCGTCACCTGCTCACGC
AGGTTCTCCCAAAAGATGGGGGCTACTATGGCAACTGCTTTTATCCAGTGAC
TGTGACAGCAATAGCTGAGGATGTTGCCACCAAAGAGTTGCTTGATGTGAT
CAAGATAATTCGGGATGGAAAGGCGAGGCTCCCCATGGAGTTTGCAAAGTG
GGCTTCAGGGGATGTGAAAGTTGATCCCTACGCATTGACATTTGAACACAA
TGTGCTTTTTGTGTCTGATTGGACGAGGTTAGGATTCTTCGAGGTAGACTAT
GGGTGGGGTACACCTAATCACATCATACCATTCACTTATGCAGACTACATGG
CAGTCGCAGTGCTTGGTGCTCCACCAATGCCAAAGAAAGGGACCCGGATTA
TGACACAGTGTGTGGAGAACAAGTGTATCAAGGAGTTCCAAGATGAGATGA
AGGCGTTCATATAA

Predicted protein sequence of *Os*FMT (*Os*AT5)

MVAVTVMRKSRNFVGPSPPTPPAEITTTLELSSIDRVPGLRHNVRSLHVFRRHK
NSGPVVDGDSRRPAAVIRAALARALADYPAFAGRFVGSLLAGDACVACTGEG
AWFVEAAADCSLDDVNGLEYPLMISEEELLPAPEDGVDPTSIPVMMQVTEFTC
GGFILGLVAVHTLADGLGAAQFITAVAELARGMDKLRVAPVWDRSLIPNPPKL
PPGPPPSFQSFGFQHFSTDVTSDRIAHVKAEYFQTFGQYCSTFDVATAKVWQAR
TRAVGYKPEIQVHVCFFANTRHLLTQVLPKDGGYYGNCFYPVTVTAIAEDVAT
KELLDVIKIIRDGKARLPMEFAKWASGDVKVDPYALTFEHNVLFVSDWTRLGF
FEVDYGWGTPNHIIPFTYADYMAVAVLGAPPMPKKGTRIMTQCVENKC
IKEFQDEMKAFI*

Text S4.1 Nucleotide sequence of *OsAT9* open reading frame

>LOC_Os01g09010.1

ATGGCGGGGACGGGGAGCTTCAAGGTGACGAGGATCTCGGAGGGGGCGGT
GAAGCCGGCGGCGGCGACGCCGGAGGAGACGCTGCCGCTGGCGTGGGTGG
ACCGGTACCCGACCCACCGCGGGCTGGTGGAGTCGATGCACATCTTCCGGT
CGGGCGCCGACGCGGCCCCGGGCGTCATCCGCGACGCGCTCGCCAGGGCGC
TGGTCTTCTTCTACCCGCTCGCCGGGCGCATCGTGGAGCCCGAGGCCGGGTC
CCCCGCCATCCGGTGCACCGCCGACGGCGTCTACTTCGCCGAGGCCGCCGC
CGACTGCAGCCTGGAGGACGTCCGCTTCCTGGAGCGCCCCCTCCTCCTCCCC
AAGGAGGACCTCGTCCCCTACCCCGGCGACGACCGCTGGGGCGTCGAGCCC
CACAACACCATCATGATGATGCAGATCACCAAATTCACCTGCGGCGGGTTC
GTGATGGGGCTCCGGTTCAACCACGCGTCGGCGGACGGCATGGGCGCGGCT
CAGTTCATCAACGCGGTGGGCGACATGGCGCGGGGGCTCCCGGAGCCGAGG
GTGAAGCCGGTGTGGGACAGGGAGAAGTTCCCGAACCCGAGCATCAAGCCC
GGCCCCCTGCCGGGGCTGCCGGTGCTGGCGCTCGACTACATCGTCCTCGACT
TCCCCACCGGCTACATCGACGGCCTGAAGGCGCAGTACAAGGCGCACAGCG
GCAAGTTCTGCTCCGGCTTCGACGTGCTGACGGCGAAGCTGTGGCAATGCC
GCACCAGGGCACTCAACCTCGAGCCCGGCGCCACCGTGAAGCTCTGCTTCTT
CGCCAGCGTGCGCCACCTGCTGAAGCTGGACAGGGGGTACTACGGCAACTC
CATCTTCCCGGTGAAGATGTCAGCGCCGAGCGAGACGGTGCTGTCGTCGTC
GGTGATGGAGGTGGTGGACATGATCCGGCAGGCGAAGGAGAGGATGGCGG
TGGAGTTCTTCCAGTTCGCCAAGGAGGAGACGGAGCAGGACCCGTTCCAGA
TGACGTTCAACTACGAGTCCATCTACGTCTCCGACTGGAGCAAGCTCGGGTT
CGCCGAGGTGGACTACGGCTTCGGCCCGCCCAAGTTCGCCGGCCCGCTCGT
CAACAACGACTTCATCGCCTCCGTCGTCATCCTCAAGGCGCCGCTGCCGCTC
GACGGCACGCGGATGCTCGCCAGCTGCGTCACCAAGGAACACTCGGAGGAG
TTCGTCCGCGGCATGAAGGAGGACCTGCCATGA

Text S4.2 Nucleotide sequence of 233-bp *OsAT9* 5'UTR selected for RNAi

AGTGCAAATCACGTTCATTGTTTCTTTTTATTTAAACTCAAATCACGTCGAG
ATTGTTGGAGAGGCGATCACACACAGCCAGCTGCATTTTTGAATTGAACGG
GTAAGCGGCATCCGTGCGCGATCTGAGCCGTTGATTCCACCGAACGCCCCC
GGCTGCTACGACACCTCCGTGTGGGGCCCACGTGACAGCGACCCACTCGCC
CAACACTCCGCCGCTCCATTTAAGGCCA